

Inter-comparison of phytoplankton functional type phenology metrics derived from ocean color algorithms and earth system models

Article

Accepted Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Kostadinov, T. S., Cabré, A., Vedantham, H., Marinov, I., Bracher, A., Brewin, R. J. W., Bricaud, A., Hirata, T., Hirawake, T., Hardman-Mountford, N. J., Mouw, C., Roy, S. and Uitz, J. (2017) Inter-comparison of phytoplankton functional type phenology metrics derived from ocean color algorithms and earth system models. *Remote Sensing of Environment*, 190. pp. 162-177. ISSN 0034-4257 Available at <https://centaur.reading.ac.uk/68387/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Inter-Comparison of Phytoplankton Functional Type Phenology

Metrics Derived from Ocean Color Algorithms and Earth System Models

Kostadinov, Tihomir S.^{1,*}, Anna Cabré^{2,**}, Harish Vedantham^{3,***}, Irina Marinov², Astrid Bracher⁴, Robert J. W. Brewin⁵, Annick Bricaud⁶, Takafumi Hirata⁷, Toru Hirawake⁸, Nick J. Hardman-Mountford⁹, Colleen Mouw^{10,****}, Shovonlal Roy¹¹, Julia Uitz⁶

[1] Department of Geography and the Environment, Univ. of Richmond, 28 Westhampton Way, Richmond, VA 23173, USA.

[2] Department of Earth & Environmental Science, Univ. of Pennsylvania, Philadelphia, PA, USA.

[3] Kapteyn Astronomical Institute, Faculty of Mathematics and Natural Sciences, Astronomy, University of Groningen, Landleven 12, 9747 AD Groningen, The Netherlands.

[4] Alfred-Wegener-Institute for Polar and Marine Research, Bussestr. 24, 27570 Bremerhaven, Germany and Institute of Environmental Physics, University Bremen, Otto Hahn Allee 1, 28359 Bremen, Germany.

[5] Plymouth Marine Laboratory (PML), Prospect Place, The Hoe, Plymouth PL1 3DH, UK.

[6] Sorbonne Universités, UPMC Univ Paris 06, CNRS, Laboratoire d'Océanographie de Villefranche, 181 chemin du Lazaret, 06230 Villefranche-sur-Mer, France.

[7] Faculty of Environmental Earth Science, Hokkaido University, N10W5, Kita-Ku, Sapporo, 060-0810 Hokkaido, Japan.

[8] Faculty of Fisheries Sciences, Hokkaido University, 3-1-1 Minato-cho, Hakodate 041-8611, Japan.

[8] CSIRO Oceans and Atmosphere Flagship, Wembley, Western Australia, Australia.

[10] Department of Geological and Mining Engineering and Sciences, Michigan Technological University, Houghton, MI, 49931, USA.

[11] Department of Geography and Environmental Science & School of Agriculture, Policy and Development,
University of Reading, Whiteknights, Reading RG6 6AB, UK.

*Corresponding author: tkostadi@richmond.edu

** Now at Dept. of Physical and Technological Oceanography, Marine Science Institute, CSIC, Barcelona, Spain.

***Now at Cahill Center for Astronomy and Astrophysics, California Institute of Technology, Pasadena, CA 91125,
USA.

****Now at University of Rhode Island, Graduate School of Oceanography, Narragansett, RI, 02882, USA

Keywords: phytoplankton bloom, phenology, phytoplankton functional types, microplankton,
ocean color algorithms, inter-comparison, CMIP5 Earth system models, Discrete Fourier
Transform.

Abstract

Ocean color remote sensing of chlorophyll concentration has revolutionized our understanding of the biology of the oceans. However, a comprehensive understanding of the structure and function of oceanic ecosystems requires the characterization of the spatio-temporal variability of various phytoplankton functional types (PFTs), which have differing biogeochemical roles. Thus, recent bio-optical algorithm developments have focused on retrieval of various PFTs. It is important to validate and inter-compare the existing PFT algorithms; however direct comparison of retrieved variables is non-trivial because in those algorithms PFTs are defined differently. Thus, it is more plausible and potentially more informative to focus on emergent properties of PFTs, such as phenology. Furthermore, ocean color satellite PFT data sets can play a pivotal role in informing and/or validating the biogeochemical routines of Earth System models. Here, the phenological characteristics of 10 PFT satellite algorithms and 7 latest-generation climate models from the Coupled Model Inter-comparison Project (CMIP5) are inter-compared as part of the International Satellite PFT Algorithm Inter-comparison Project. The comparison is based on monthly satellite data (mostly SeaWiFS) for the 2003-2007 period. The phenological analysis is based on the fraction of microplankton or a similar variable for the satellite algorithms and on the carbon biomass due to diatoms for the climate models. The seasonal cycle is estimated on a per-pixel basis as a sum of sinusoidal harmonics, derived from the Discrete Fourier Transform of the variable time series. Peak analysis is then applied to the estimated seasonal signal and the following phenological parameters are quantified for each satellite algorithm and climate model: seasonal amplitude, percent seasonal variance, month of maximum, and bloom duration. Secondary/double blooms occur in many areas and are also quantified. The algorithms and the models are quantitatively compared based on these emergent phenological parameters. Results

indicate that while algorithms agree to a first order on a global scale, large differences among them exist; differences are analyzed in detail for two Longhurst regions in the North Atlantic: North Atlantic Drift Region (NADR) and North Atlantic Subtropical Gyre West (NASW). Seasonal cycles explain the most variance in zonal bands in the seasonally-stratified subtropics at about 30° latitude in the satellite PFT data. The CMIP5 models do not reproduce this pattern, exhibiting higher seasonality in mid and high-latitudes and generally much more spatially homogeneous patterns in phenological indices compared to satellite data. Satellite data indicate a complex structure of double blooms in the Equatorial region and mid-latitudes, and single blooms on the poleward edges of the subtropical gyres. In contrast, the CMIP5 models show single annual blooms over most of the ocean except for the Equatorial band and Arabian Sea.

1. Introduction

Marine phytoplankton play an important role in the global carbon cycle via oxygenic photosynthesis and the biological pump (*Field et al., 1998; Eppley and Peterson, 1979; Falkowski et al., 1998; IPCC, 2013; Siegel et al., 2014*). Since the late 1990's, ocean color remote sensing has enhanced our understanding of oceanic ecosystems via continuous global estimates of total chlorophyll *a* concentration (Chl), interpreted as a proxy for phytoplankton biomass (e.g. *McClain, 2009; Siegel et al., 2013*). However, total Chl does not provide a full description of the ecosystem. Phytoplankton have different morphological (size and shape) and physiological (growth and mortality rates, response to nutrient, temperature and light conditions) characteristics and different resulting biogeochemical and ecological roles (e.g. silica or iron requirements, calcification, sinking rates, feeding characteristics) and are thus grouped accordingly into phytoplankton functional types (PFTs, e.g. *IOCCG 2014*). Phytoplankton community structure influences many fundamental components of the marine biogeochemical cycle, including: phytoplankton physiology; nutrient uptake; nutrient cycling; growth rates; metabolic rates; deep-ocean carbon export; and the transfer of energy through the marine food web (*IOCCG, 2014*). Therefore, detailed characterization of PFTs, and not only total Chl, is required to develop predictive understanding of the ocean's role in climate on various time scales (e.g. *Le Quéré et al., 2005; Hood et al., 2006; Stock et al., 2014*) and inform climate models. One of the primary distinguishing characteristics of the different PFTs is cell size, which is considered to be a master trait (*Marañón, 2015*) and is correlated to first order with biogeochemical function – e.g. *Le Quéré et al., 2005*). Size partitioning has been used as a first-order proxy for PFT classification (e.g. *Vidussi et al., 2001; Le Quéré et al., 2005; Uitz et al., 2006; Kostadinov et al., 2010*).

Satellite remote sensing provides a comprehensive observation method to characterize the global spatio-temporal distribution of PFTs (e.g. *McClain et al. 2009; Siegel et al. 2013*). Space-borne platforms can provide continuous sampling at the required resolution in time and space in order to facilitate the development of more complex “dynamic green ocean models” (*Le Quéré et al., 2005*) that include multiple functional types and resolve important biogeochemical processes (*IOCCG, 2014, Ch. 1, Sect. 1.5*). Multiple satellite bio-optical algorithms for the retrievals of various PFTs have been developed in the last decade as a result. One class of algorithms is based on total abundance and the premise that smaller cells are associated with oligotrophic conditions whereas larger cells are associated with eutrophic conditions (*Chisholm, 1992*) – such algorithms are described by *Brewin et al. (2010)*, *Hirata et al. (2011)* and *Uitz et al. (2006)*. Another class of algorithms relies on various spectral features. The PHYSAT algorithm exploits second-order anomalies of reflectance spectra (*Alvain et al., 2005; Alvain et al., 2008*), whereas several other algorithms are based on either absorption (*Bracher et al., 2009; Ciotti and Bricaud, 2006; Mouw and Yoder, 2010; Roy et al. 2011; Roy et al 2013*), or backscattering (*Kostadinov et al., 2009; Kostadinov et al., 2010; Kostadinov et al., 2016*), or a hybrid of absorption and backscattering (*Fujiwara et al., 2011*).

Brewin et al. (2011) conducted the first systematic inter-comparison of PFT algorithms designed to identify “dominant” PFTs in the oceans. With the increasing publication of new PFT algorithms (*IOCCG, 2014*), an international team of PFT algorithm developers and scientists was tasked to perform a follow-up inter-comparison exercise (*Hirata et al., 2012; Hirata, 2015*); this study reports results from a component of this inter-comparison project. A summary of the

138 available algorithms and their technical basis can be found in Table 1 (also see *IOCCG, 2014*).
139 The various algorithms use different PFT definitions and retrieve different variables that are
140 based on various sets of assumptions, and hence are not necessarily directly comparable. Some
141 retrieve several taxonomic groups, others – size fractions based on Chl or volume (Table 1;
142 *IOCCG 2014*). PFT algorithms often aim to quantify the size structure of the phytoplankton
143 population by defining three phytoplankton size classes (PSCs) – picoplankton ($< 2 \mu\text{m}$),
144 nanoplankton ($2\text{-}20 \mu\text{m}$), and microplankton ($> 20 \mu\text{m}$) (*Sieburth et al., 1978*). This is justified
145 because size is considered a master trait (e.g. *Marañon et al. 2015*), but we caution that
146 differences exist between PFTs and PSCs, even though this terminology is often used
147 interchangeably.

Table 1. Overview of the PFT/PSC algorithms used and the relevant variable(s) from which phenological parameters were derived. SW10 refers to SeaWiFS monthly mapped 9 km global $R_{rs}(\lambda)$ data for the 2003-2007 period. Monthly data for 2003-2007 from SCIAMACHY on ENVISAT was only used for PhytoDOAS and has 1/2 degree spatial resolution. The variables provided by most algorithm are dimensionless, i.e. fractions of a total, most commonly – chlorophyll-a (Chl). This is indicated by a double dash in the table. If *in situ* data were used in algorithm development, the region from which the data came is indicated. N/A means no *in situ* data were directly used in the algorithm development (not including validation) (see references for details).

Algorithm Publication(s)	Acronym	Variables Analyzed	Units	Input Data	Algorithm Class/Basis	Variables Retrieved	Region of development
<i>Alvain et al. (2005,2008)</i>	PHYSAT	Frequency of detection of diatoms	% of days	SW10	$R_{rs}(\lambda)$ second-order anomalies (Radiance-based)	Multiple taxonomic PFTs	North Atlantic; Equatorial & Tropical South Pacific; Southern Ocean
<i>Bracher et al. (2009); Sadeghi et al. (2012)</i>	PhytoDOAS	Diatoms Chl	mg m ⁻³	SCIAMACHY	Differential absorption from hyperspectral data. (Absorption-based)	Multiple taxonomic PFTs	Uses PFT-specific $a_{ph}(\lambda)$
<i>Brewin et al. (2010)</i>	BR10	Microplankton – fraction of Chl	--	SW10	Abundance-based.	Size structure	Atlantic Ocean
<i>Ciotti and Bricaud (2006), Bricaud et al. (2012)</i>	CB06	$1 - S_f$ where S_f = fraction of small phytoplankton	--	SW10	Absorption-based.	Size structure	Global (<i>In situ</i> data used for picoplankton basis vector)
<i>Fujiwara et al. (2011)</i>	FUJI11	Microplankton – fraction of Chl	--	SW10	Absorption- and backscattering-based.	Size structure	Arctic-North Pacific
<i>Hirata et al. (2011)</i>	OC-PFT	Microplankton –	--	SW10	Abundance-based.	Size structure	Global (coastal and

		fraction of Chl					shelf waters excluded)
<i>Kostadinov et al. (2009, 2010)</i>	KSM09	Microplankton - volume fraction	--	SW10	Backscattering-based.	Size structure	N/A
<i>Mouw and Yoder (2010)</i>	MY10	S_{fm} , fraction of large phytoplankton	--	SW10	Absorption-based.	Size structure	Global
<i>Roy et al. (2011, 2013)</i>	ROY13	Microplankton – fraction of Chl	--	SW10	Absorption-based.	Size structure	Global
<i>Uitz et al. (2006)</i>	UITZ06	Microplankton – fraction of Chl	--	SW10	Abundance-based.	Size structure	Global (case-2 waters excluded)
<i>O'Reilly et al. (1998, 2000)</i>	OC4v6	Chl-a	mg m ⁻³	SW10	Band-ratio $R_{rs}(\lambda)$ based. (Radiance-based)	Chl-a	Global

Here we compare the algorithm outputs in terms of a key emergent property: phytoplankton phenology. Since seasonal cycles are a key property of ecosystems, it is important to assess to what degree different algorithms agree in terms of phenology, i.e. how consistently they capture the annual progression of phytoplankton blooming and subsequent senescence. If the timing of a bloom were slightly shifted between two data sets, direct comparison of the variables at each time step would yield disparate and meaningless results, whereas phenological analysis will identify the offset in timing (*Platt et al., 2009*). The amount of algorithms spread about an ensemble mean can be indicative of our confidence in retrieving a certain phenological parameter (e.g. timing of annual bloom), and overall results of the comparison can guide further algorithm improvements.

We use the Discrete Fourier Transform (DFT) to first model the seasonal cycle as a summation of sinusoids derived from the annual frequency band and its harmonics (integer multiples). We then quantify phenological parameters of interest using the modeled seasonal cycle. The phenology inter-comparison is based on global ocean color data (SeaWiFS and SCIAMACHY) for the period 2003-2007, using microplankton fraction or the most similar available variable from each participating PFT/PSC algorithm (Table 1). Increases in the absolute or fractional amount of large phytoplankton or diatoms were considered here to define a bloom, which is consistent with the established ecological idea that higher chlorophyll concentrations are associated with eutrophy and a relatively higher dominance of large phytoplankton (e.g. *Chisholm, 1992; Loisel et al., 2006; Kostadinov et al., 2010; Marañón, 2015*). We quantify the timing, amplitude and duration of blooms, as well as the fraction of variance explained by the modeled seasonal cycle. We compare these phenological parameters among the PFT algorithms. The same phenological

parameters are also compared for the NASA chlorophyll product (OC4v6 Chl), as well as contemporary diatom carbon biomass provided by seven CMIP5 Earth System Models (ESMs). Our goal is not to rank the satellite algorithms and CMIP5 models in terms of quality; rather we strive to identify spatial patterns of agreement and disagreement among the algorithms in an effort to guide future improvements. Additionally, the comparison to the ESM ensemble is aimed at guiding future improvements in biogeochemical and climate modeling, a key goal of the Earth system science community (*IPCC, 2013*).

2. Data and Methods

2.1 Input Satellite Data

All algorithms with the exception of PhytoDOAS use monthly global 9 km Level 3 mapped SeaWiFS remote-sensing reflectance, ($R_{rs}(\lambda)$), reprocessing R2010.0) from January 2003 to December 2007 as input (60 monthly maps total). These data as well as the corresponding monthly OC4v6 Chl data (*O'Reilly et al., 2000*) and monthly composites of daily averaged photosynthetically available radiation (PAR, mol photons m⁻² day⁻¹) from the same reprocessing were downloaded from the NASA Ocean Biology Distributed Active Archive Center (OB.DAAC) maintained by the Ocean Biology Processing Group (OBPG) (<http://oceandata.sci.gsfc.nasa.gov/>) (*NASA Goddard Spaceflight Center, 2010*). $R_{rs}(\lambda)$ data were processed by the individual algorithm providers. The hyperspectral PhytoDOAS algorithm is based on Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) level-1 top-of-atmosphere radiance data. SCIAMACHY was a satellite sensor with a native pixel size is 30 km by 60 km which operated from 2002 to 2012 on the ENVISAT satellite. These processing details were agreed upon by the International PFT Inter-comparison

Project Team (*Hirata et al., 2012*). SeaWiFS Chl data were analyzed in the same way as the PFT algorithms data for comparison purposes. PAR data were used for verification of the DFT phenological algorithm (Supplement Part 1).

2.2 PFT/PSC Algorithm Output Pre-Processing

The PFT/PSC algorithms were used to derive phenological parameters using the variable most closely corresponding to either microplankton Chl [mg m^{-3}] or microplankton/large phytoplankton fraction [% of total Chl]. The rationale behind this choice is 1) blooming/more eutrophic conditions are on average characterized by an increase in total and fractional large phytoplankton biomass; and 2) this is the most common variable among all available algorithms. Table 1 summarizes the respective variables used in the phenological analysis, indicates the acronym used here for each algorithm and provides additional relevant information. For additional algorithm methodologies details, see the references in Table 1 and *IOCCG (2014)*.

Monthly data from all algorithms were down-sampled to 1 degree resolution using two-dimensional convolution with a 12 x 12 top hat averaging kernel (2 x 2 in the case of PhytoDOAS due to its different resolution). Missing data in the original resolution were ignored in the averaging; however, if less than 50% of the pixels being averaged were valid data, the pixel in the down-sampled image was assigned a missing data value. For PhytoDOAS, even a single valid pixel of the four being averaged produced a valid pixel in the down-sampled image. The registration of PhytoDOAS images was changed from grid/node to cell/pixel (*NCEI, 2015*) in order to match all other down-sampled imagery. All variables were down-sampled in linear space with the exception of Chl and the PhytoDOAS data, which were down-sampled in log10 space,

since Chl values tend to vary lognormally spatially (*Campbell, 1995*). The log space spatial average was weighted appropriately for any present zeros, which cannot participate in a log average (*Habib, 2012*). Note that taking an arithmetic average in log space approximates the median of the data in linear space (*Campbell, 1995*).

2.3 Phenological Parameters via Discrete Fourier Transform (DFT); Metrics of Algorithm and Model Inter-comparison.

An increase in the absolute or fractional amount of large phytoplankton or diatoms is considered a bloom, and the bloom peak is considered the maximum of these values, respectively. A time series of each algorithm's relevant microplankton or diatom variable (Table 1) was constructed at each pixel at 1-degree resolution. Data were gap-filled temporally by linear interpolation (no extrapolation was applied). If more than 45% of the data points were missing or if there was a continuous run of missing data longer than 8 months anywhere in the time series, data for that pixel was not used in the analysis. The mean was subtracted from each time series. Interannual variability is not explicitly considered in this study; however, possible secular trends are removed by detrending (by subtracting a least-squares line fit to the data), and other interannual variability in the study period of 2003-2007 is taken into account implicitly because the DFT is computed over the entire time series. If data were missing at the edges of the time series, they were filled with zeros. The DFT was then used to transform the time series to the frequency domain. The Fourier coefficients at a frequency of one cycle per year ($f = 1 \text{ yr}^{-1}$) and all its available harmonics (integer multiples) were used to model the seasonal cycle as a sum of sinusoids of varying phases and amplitudes. Supplement Part 1 describes the details of the DFT analysis and seasonal cycle modeling, and Supplement Fig. S1 illustrates an example modeled seasonal cycle.

248

249 Peak analysis was performed on the modeled annual cycle signal using the MATLAB® (R2014b)
250 Signal Processing Toolbox® routine `findpeaks` in order to determine the timing of the local
251 maxima, the signal height (the value of the signal at the peak) and the width of the signal at half-
252 height. The modeled signal minimum value was subtracted from the signal before peak analysis
253 to ensure correct height determinations. Signal edge effects were taken into account. In order to
254 avoid detection of small secondary peaks (many of which can be artifacts of the modeling), only
255 peaks whose prominence was greater than 10% of the signal range (maximum minus minimum
256 value) and which were at least 2.5 months apart from each other were detected. Prominence here
257 is equivalent to topographic prominence and can be thought of as the intrinsic height of the peak
258 relative to other nearby peaks. The same phenological analysis was applied to diatom carbon
259 output from 7 CMIP5 models – details of the methodology and model information are provided
260 in Supplement Part 2 and in *Cabré et al. (2015)*. The following phenological parameters and
261 inter-comparison metrics were derived from the peak analysis for both the PFT algorithms and
262 the CMIP5 models:

263

264 1) ***Seasonal amplitude*** of the primary bloom, determined as half the height of the most prominent
265 (highest) peak (Supplement Fig. S1), was inter-compared qualitatively because variables are on
266 different scales (even among algorithms that have the same units, the methodologies are
267 different).

268 2) The ***month of maximum*** of the primary bloom, determined as the month where the DFT-
269 modeled seasonal cycle is maximum. The ensemble mean for the algorithms and CMIP5 models
270 (calculated separately for each ensemble) was used to quantitatively compare the month of

maximum of each algorithm to the ensemble mean month of maximum for all algorithms. We also compared the Chl and the ensemble mean CMIP5 model months of maxima. Variances in month of maxima were quantified for the algorithms and the CMIP5 models separately. The month of maximum was averaged across algorithms and models using circular statistical methods to ensure a properly estimated mean and variance (Supplement Part 3).

3) **Duration** of the primary bloom (in days), determined as the width of the most prominent peak at the peak's half-height level. Ensemble mean and standard deviation of bloom duration were calculated for the CMIP5 models and algorithms, and individual algorithm durations were qualitatively compared.

4) **Percent seasonal variance**, i.e. fraction of the data variance explained by the modeled seasonal cycle as opposed to other processes (e.g. one-time events, multiannual processes, and inter-annual variability due to climate oscillation modes like ENSO) and noise. An area in which this fraction is very high is characterized by a very clean seasonal signal i.e. little variance contribution by other processes. It is calculated as the sum of power at $f = 1 \text{ yr}^{-1}$ and its harmonic frequencies (Supplement Part 1), divided by total variance of the input data. The ensemble mean among the algorithms and models was calculated.

At least three algorithms (or two CMIP5 models) were required to participate in the ensemble means of month of maximum, percent seasonal variance and bloom duration for the ensemble statistics to be considered valid. If present, the second most prominent peak, representing a possible secondary bloom, was also characterized by using the above phenological parameters and the following was also derived: difference in months between the primary and secondary bloom, and relative prominence of the secondary blooms (ratio of the prominence of the

secondary peak to the prominence of the primary peak). The fraction of satellite algorithms or CMIP5 models exhibiting two annual peaks was mapped, indicating in which regions the majority of algorithms or models agree that there is a secondary bloom. The difference in months between the maxima of the primary and secondary blooms was compared qualitatively among the algorithms. Finally, regionally binned analysis was performed for the following *Longhurst (1998)* provinces 1) Westerlies - North Atlantic Drift – NADR and 2) North Atlantic Subtropical Gyre West - NASW. Available data in these regions were spatially averaged and the resulting single time series per region and algorithm/model were analyzed. Methodological details of this analysis are provided in Supplement Part 4.

3. Results and Discussion

3.1 Seasonal Amplitude

The seasonal amplitude (Fig. 1) of the relevant PFT variables (Table 1) quantifies the strength of the seasonal cycle at a particular location. The tropical ocean and the oligotrophic subtropical gyres (defined here as the regions delineated by the climatological SeaWiFS Chl = 0.08 mg m⁻³ isoline and having Chl values less than this value) were generally characterized by low seasonal amplitudes of Chl and microplankton across all the algorithms. However, according to most algorithms, the southern edge of the South Pacific gyre, at around 30°S, was characterized by a band of higher seasonal amplitude, particularly evident in the UITZ06 and ROY13 data. Notably, in BR10 and UITZ06, some of the areas of lowest amplitude for microplankton are just equatorward of the Chl isoline delineating the gyre. Similarly, in ROY13 the gyres are not actually the places with the lowest amplitudes; rather they occur just equatorward of them in the Pacific. According to most algorithms, the highest seasonal amplitudes occur at the temperate

316 and subpolar latitudes and in coastal zones, particularly the North Atlantic, the Northwestern
317 Pacific, and in the monsoon-driven upwelling region off of the Arabian Peninsula. Circumpolar
318 bands of high and low seasonal amplitude characterize the Southern Ocean, but the spatial details
319 of the bands differ across algorithms. Overall, the large-scale spatial patterns of seasonal
320 amplitude are similar, across algorithms of the same type. The CB06 algorithm retrieved almost
321 no valid data over the gyres, and the PHYSAT data sparsity precludes meaningful DFT analysis
322 over most areas; thus no PHYSAT map was included in Fig. 1 and subsequent maps (See Sect 3.6
323 for regionally binned analyses of PHYSAT data).

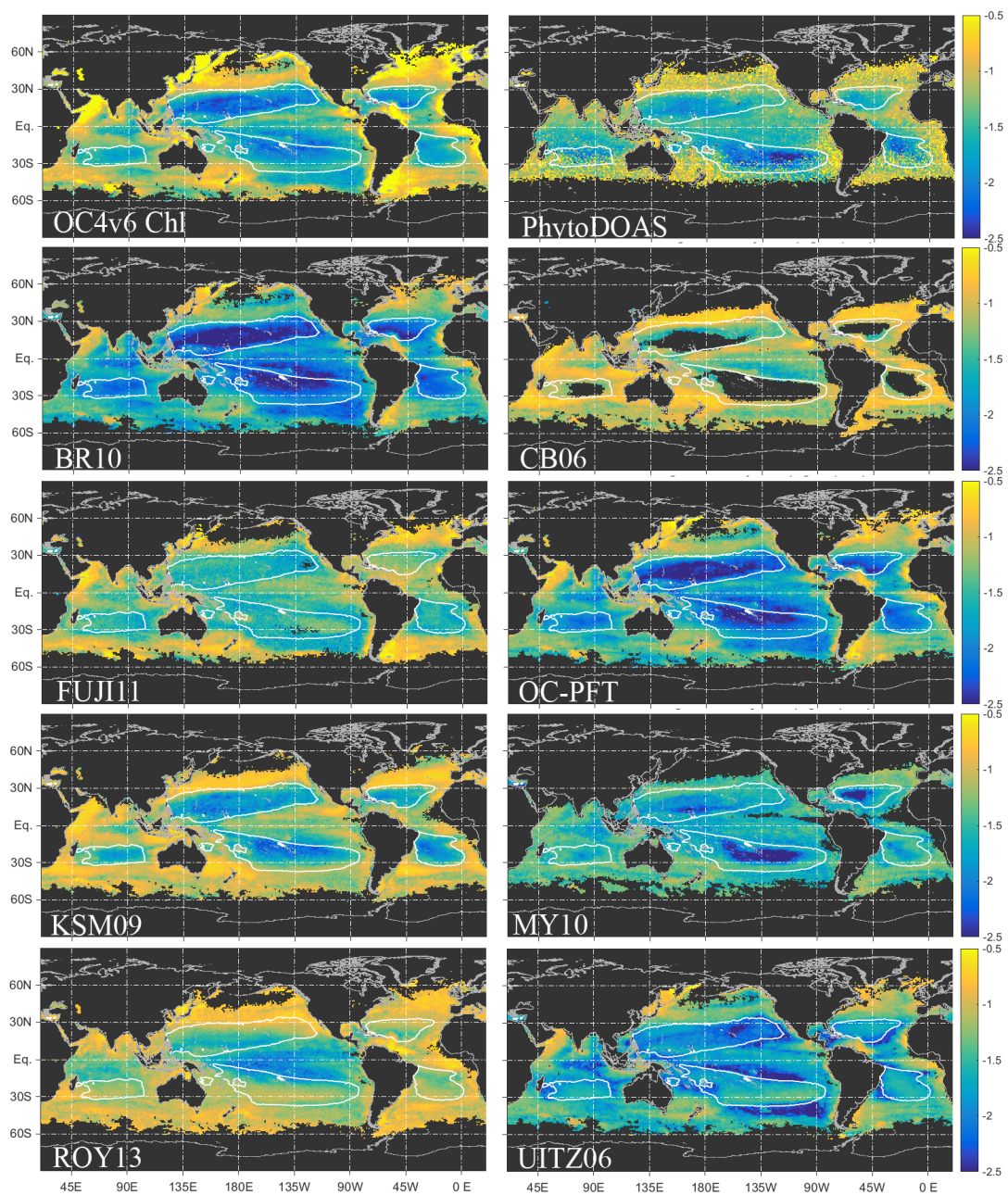


Figure 1. Seasonal amplitude of Chl and the large phytoplankton/diatoms variables of 9 PFT satellite ocean color algorithms (Table 1). PHYSAT is not shown due to data sparsity. The same logarithmic color scale applies to all maps, and units are as

indicated in Table 1. The isoline of climatological Chl = 0.08 mg m^{-3} is plotted as a solid white contour. All pixels where valid phenological analysis was performed are mapped.

3.2 Percent Seasonal Variance

The ensemble-mean percent seasonal variance for all 10 PFT algorithms (Fig. 2A) reveals several oceanic zones where the seasonal cycle is particularly clean/reproducible. This is especially prominent at the poleward boundaries of the subtropical gyres, around 30-40° latitude in both hemispheres, where 70-80% of the signal variability is explained by the seasonal cycle represented by the DFT. These highly-seasonally variable regions correspond to sharp transitions in surface Chl concentration, as observed in *in situ* observations and satellite data (e.g. *Glover et al. 1994*). In the Pacific this feature is known as the transition zone chlorophyll front (TZCF), which migrates from 30-35°N in winter to 40-45°N in summer; its migration is due primarily to wind-driven seasonal variations in Ekman pumping and Ekman advection of nutrients (e.g., *Bograd et al. 2004*). The equivalent North Atlantic seasonally-stratified subtropics coincide with the mid-latitude biome of *Levy et al. (2005)* and are described as a nutrient-limited regime. Equivalent seasonally-stratified, nutrient-limited subtropics with high seasonal variability are present in the Southern Ocean band around 30°S.

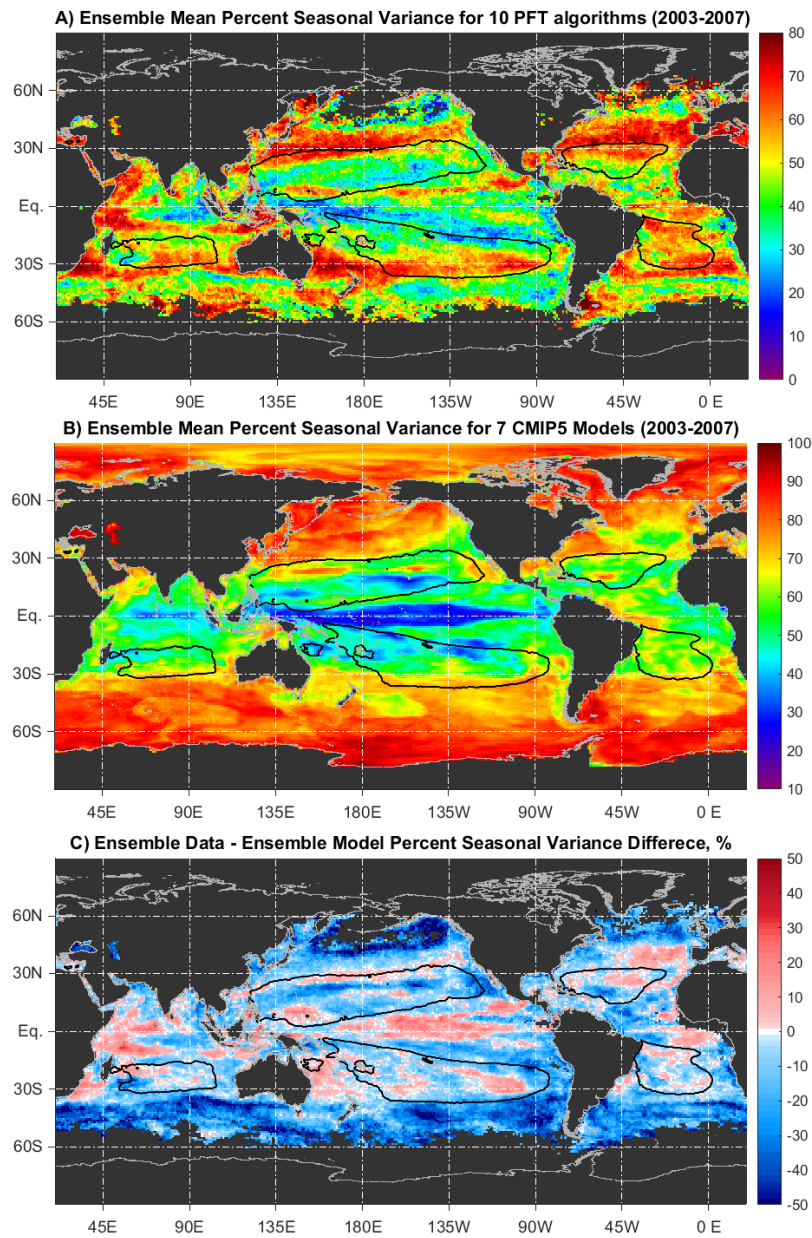


Figure 2. Ensemble mean percent seasonal variance for the 2003-2007 period for A)

the 10 PFT algorithms (Table 1) and B) the 7 CMIP5 models (Table S1). (C) The

difference in percent seasonal variance between the satellite data and the models

(positive difference means satellite data percent seasonal variance is larger than the

model value). The isoline of climatological Chl = 0.08 mg m⁻³ is plotted as a solid black

contour.

Detailed analysis of the KSM09-based carbon biomass from SeaWiFS (*Kostadinov et al., 2016*) and CMIP5 model output shows that the regions of high percent seasonal variance of Fig. 2A exhibit strong and reproducible seasonality and are mainly dominated by nano- and microphytoplankton during the bloom months, and by picophytoplankton during the low-biomass summer months (*Cabré et al. 2016*). These ensemble mean PFT-based results are consistent with the Chl-based analysis of *Sapiano et al. (2012)* who also determined that the poleward fringes of the subtropical gyres have the best seasonality statistical fits

In contrast to the satellite data, CMIP5 model diatom biomass exhibits smoother spatial variability of percent variance explained by the seasonal cycle (Fig. 2B, model ensemble mean; Fig 2C, model-data difference map), with much broader regions characterized by more than 60% of variance due to the annual cycle. The percent variance explained by the seasonal cycle is a much stronger function of latitude (and hence the seasonality of insolation) in models. Poleward of 40°S/40°N, models show stronger seasonal variance compared to satellite data. In contrast, the Equatorial regions, especially in the Pacific and Indian Oceans, have lower percent annual variance as compared to the satellite data.

3.3 Month of Maximum of the Primary Bloom

The ensemble-mean month of maximum of the primary peak (Fig. 3A) varies as a function of latitude to first order, due to the seasonality of insolation, which is a primary physical driver of oceanic ecosystems, controlling both light and nutrient availability (via heating rates, wind patterns and mixing). This latitudinal dependence is most pronounced in the North Atlantic,

378 where the subtropics experience a late winter-early spring bloom, northern temperate latitudes
379 experience maximum blooms in May and June, and subarctic regions – as late as August.
380 Equivalently, for much of the Southern Hemisphere subtropical seas, the maximum of the bloom
381 occurs in late austral winter – July and August (Fig. 3A). The Southern Ocean blooms later
382 during austral late spring and summer, mostly in November through February. However, the
383 Southern Ocean exhibits an interesting banded structure where large phytoplankton
384 (microplankton, diatoms) bloom earlier (November and December) in a zonal band around 50°S,
385 as compared to a nearly continuous band just to the north, at about 45°S, which tends to bloom
386 later in January and February.

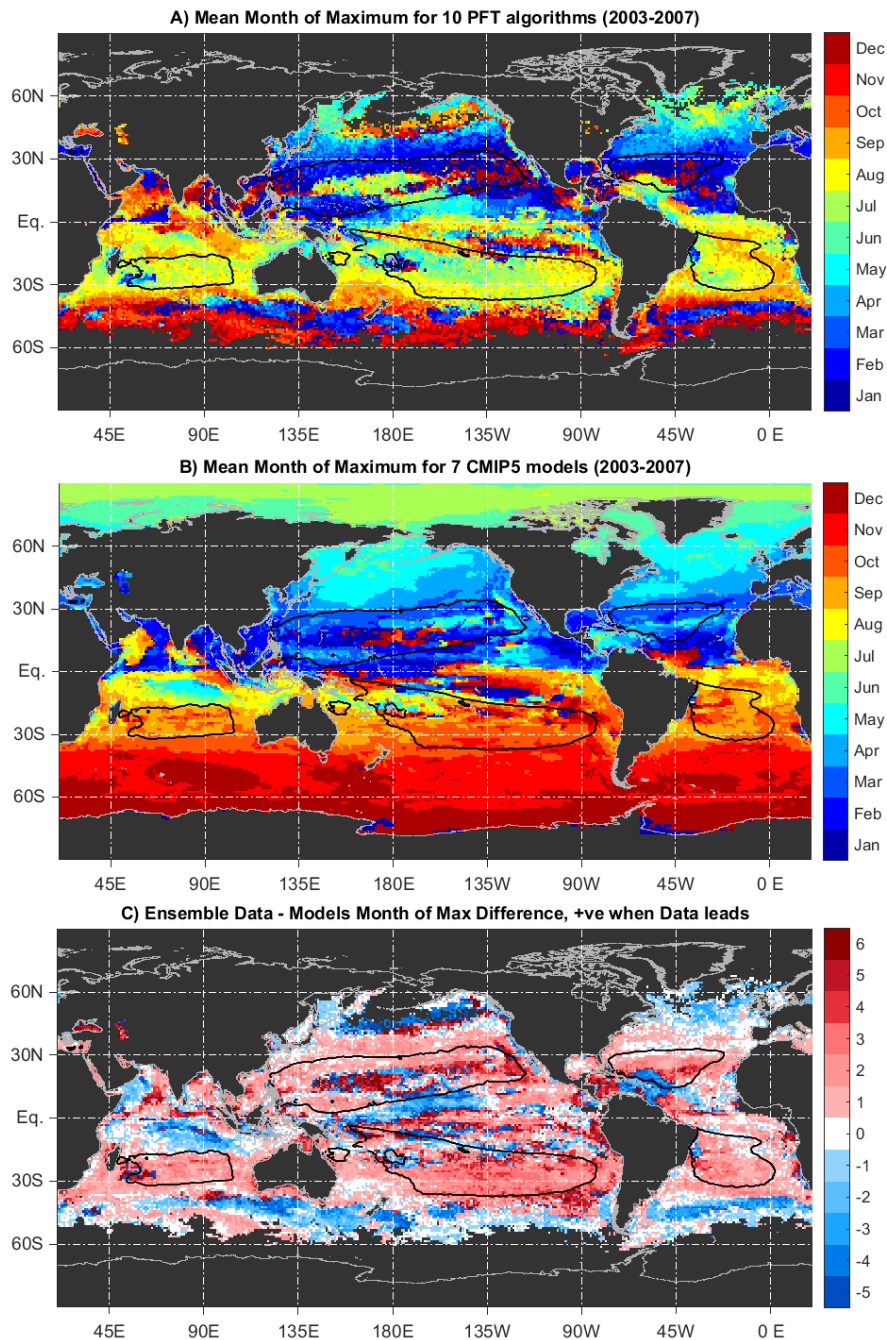


Figure 3. Ensemble mean month of maximum of A) fraction large phytoplankton/diatoms (or Chl) among 10 PFT satellite algorithms (Table 1), and B) diatom biomass among 7 CMIP5 models. C) The difference (in months) between the ensemble means of the satellite data and the models. A positive difference means the satellite data is leading

the models, i.e. the bloom peak occurs earlier in the data than in the model. The isoline of climatological Chl = 0.08 mg m⁻³ is shown (black solid contour). The means and the respective difference should be treated with caution in areas where a considerable number of algorithms or models exhibit low fraction of variance explained by the seasonal cycle (See. Supplement Fig. S13).

The CMIP5 ensemble mean month of maximum of diatom carbon biomass (Fig. 3B) exhibits a similar spatial pattern with latitude and is spatially less noisy than the satellite PFT estimates. In general the models place the blooms later in time over most of the ocean, as indicated by the algorithm-model difference map (Fig. 3C, red colors indicate the data peak leads the model peak in time). Notable exceptions are some areas in or near the subtropical gyres, the Equatorial Upwelling, and the higher latitudes (e.g. the models do not reproduce the aforementioned banded structure in the Southern Ocean), where models place the blooms earlier in time. The algorithm-model difference is about one month over much of the ocean area (Fig. 3C, pale red or blue). This difference is not randomly distributed and exhibits definite spatial patterns, pointing to latitudinal biases in processes and understanding of seasonality in models.

The differences between the PFT algorithm's ensemble mean month of maximum and the month of maximum for Chl are small (Supplement Fig. S2, top left panel), indicating that the PFT algorithm ensemble mean month of maximum for microplankton (or similar variable, Table 1) appears to be representative of that for total Chl. Several algorithms (BR10, OC-PFT, UITZ06) are abundance-based (Table 1), i.e. their PFT retrievals are a strong function of Chl; thus it is not surprising that their individual differences with the Chl month of maximum are relatively small.

So the ensemble mean month of maximum may be driven by the abundance-based PFT algorithms. Two of the spectral-based models (CB06 and MY10) also exhibit generally small differences with the Chl results. Differences among other algorithms with respect to the month of maximum can be larger; while for most of the ocean and for most algorithms the differences are not very large, considerable discrepancies persist in significant ocean areas where month of maxima difference can reach up to 5 or 6 months. More details, including possible reasons for the observed differences, are discussed in Supplement Part 5. An alternative way to quantify the level of agreement among the satellite algorithms or the CMIP5 models is the circular variance of the month of maxima (Supplement Fig. S3). Note that in areas where percent seasonal variance is low (Fig. 2A and Supplement Fig. S13), the concept of month of maximum for the seasonal cycle breaks down and results in these areas should be interpreted with caution. See Sect. 3.7 and this Supplement Part 9 for details.

We note that from a resource management standpoint, a difference of a month can be very significant, especially with respect to the effect on higher trophic levels. For example, *Platt et al. (2003)* conclude that differences of three weeks in the timing of the spring algal bloom can have large influences on the survival index of fish larvae in the Northwest Atlantic. *Koeller et al. (2009)* discuss the coupling of the phenologies of phytoplankton and shrimp in the North Atlantic. In general, whether a difference of one or two months among the algorithms and CMIP5 models is significant for practical applications will depend on the specific application. Which PFT algorithm or algorithm ensemble may be best suited to inform a certain decision will also depend on the issue at hand, as the algorithms have different theoretical bases. However, in

general, practical applications would best be addressed by using daily or 8-day data, rather than the monthly data used here for the global inter-comparison.

3.4 Primary Bloom Duration

The ensemble mean of the duration of the primary annual bloom among the 10 PFT algorithms (Fig. 4A) indicates that over much of the ocean the bloom duration is about 100 – 120 days, i.e. about 3-4 months. Maximum durations tend to occur at the poleward and to a lesser extent, the equatorward fringes of the subtropical gyres. These maximum bands are most prominent in the Pacific. Large portions of the interior of the southern hemisphere subtropical gyres also exhibit long bloom duration. These results are consistent to first order with the SeaWiFS Chl-based phenological analysis of *Racault et al. (2012)* and *Sapiano et al. (2012)*, who use different methodologies. *Sapiano et al. (2012)* observe longer bloom durations than the analysis here. They note that their results are indeed longer than most previous studies and also caution that in areas of double blooms, their duration indicates the combined duration of the blooms in some cases. Additionally, our analysis may exhibit shorter durations if there is a taxonomical succession, as our analysis indicates the bloom of only microplankton/diatoms. *Sapiano et al. (2012)* note that bloom durations do not tend to exhibit a simple pattern of decrease with higher latitudes, which is consistent with our observations (Fig. 4A), and different from the result of *Racault et al (2012)*. The PFT ensemble mean exhibits high spatial frequency noise, and there is no clear pattern of decreasing bloom duration with increasing latitude. The same is noted by *Sapiano et al. (2012)* and is also apparent in the analysis of *Racault et al. (2012)* to some degree, but note that they use a coarser spatial smoothing. Maps of primary bloom duration for Chl and the individual PFT algorithms are shown in Supplement Fig. S4 and agreement among the

algorithms and CMIP5 models is quantified by the variance in primary bloom duration
(Supplement Fig. S5); additional details are discussed in Supplement Part 6.

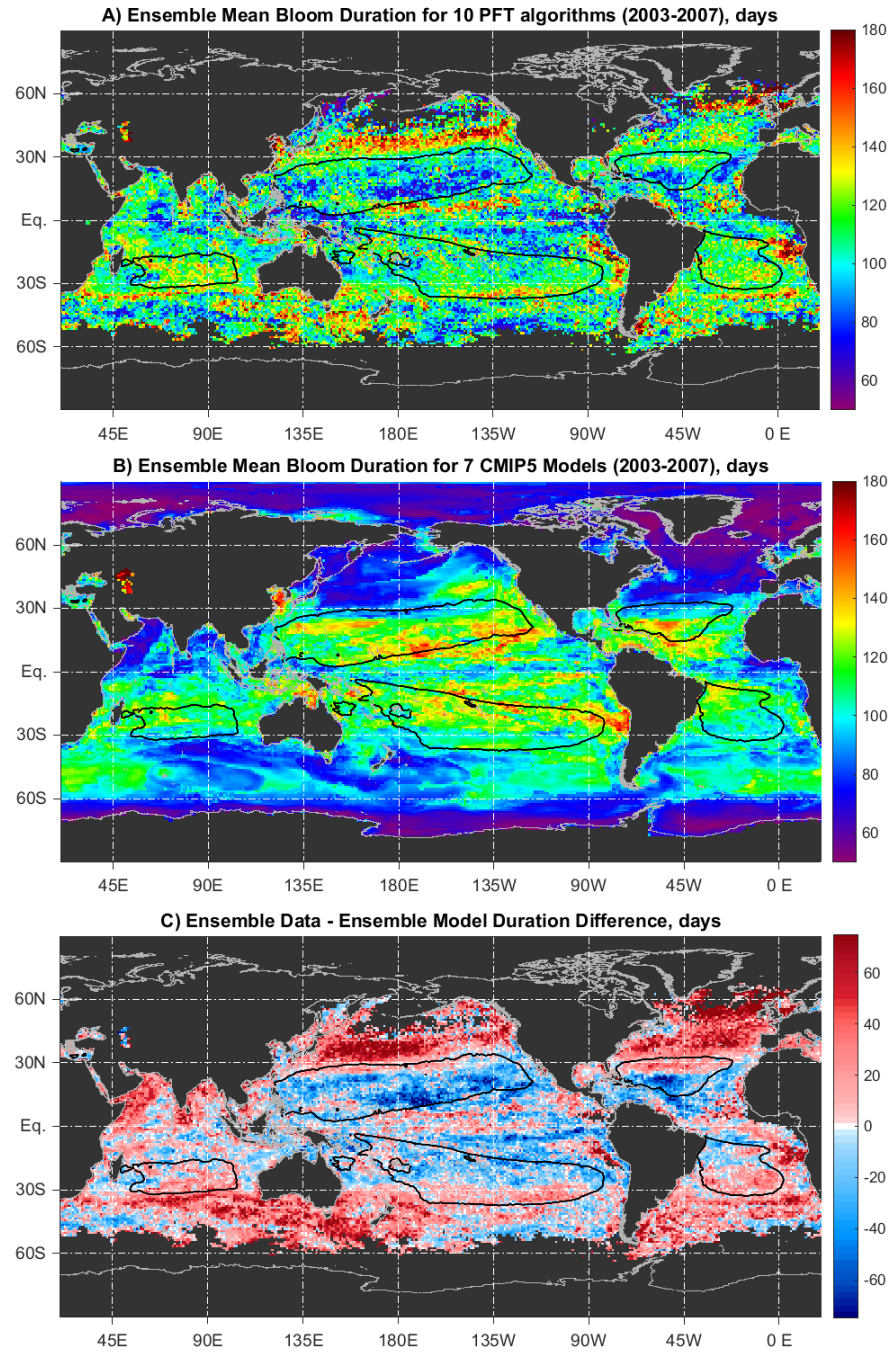


Figure 4. Ensemble mean primary bloom duration (in days) for A) large phytoplankton/diatoms among 10 PFT satellite algorithms, and B) diatom biomass among 7 CMIP5 models. C) Difference in bloom duration between the satellite data and the models (positive when data bloom duration is larger). Duration is defined as the width of the modeled seasonal signal at half the bloom peak height. The isoline of climatological Chl = 0.08 mg m⁻³ is shown (black solid contour).

The CMIP5 models, in contrast to the PFT algorithm data, exhibit much smoother bloom duration spatially (Fig. 4B), and the most prominent duration maxima occur at the equatorward fringes and inside of the subtropical gyres. Compared to the satellite data, the higher latitudes exhibit a much more obvious progression towards shorter bloom durations, dropping to below 2 months for polar latitudes. Models fail to capture secondary peaks, which are especially important at high latitudes, as explained in the following section. The lack of secondary peaks might contribute to a shorter and cleaner definition of bloom duration in models when compared to data. The difference in bloom duration between the models and the data (Fig. 4C) confirm that in general, models exhibit longer blooms in the gyres and shorter bloom at latitudes higher than ~ 30°.

3.5 Secondary Blooms

The presence of secondary blooms can be detected, because several harmonics were used in the DFT analysis (Sect. 2.3 and Supplement Sect. S1). In the mid-latitudes, the second bloom is usually a secondary bloom of smaller amplitude in the respective hemisphere's autumn (e.g. *Sapiano et al, 2012*). To summarize the PFT algorithm and model consensus about where secondary blooms occur, the fraction of algorithms that exhibit a single annual peak (Fig. 5A) versus a double annual peak (Fig. 5B) is employed. Both maps exhibit well-defined latitudinal

banding, where most algorithms exhibit a single peak at the poleward fringes of the subtropical gyres ($\sim 30^\circ$ latitude), a double peak around $40\text{--}45^\circ$ in both hemispheres, and again a single peak at higher sub-polar latitudes of about 60° . Previous studies have identified and studied this pattern, using Chl data (*Sapiano et al., 2012*) together with ecosystem modeling (*Platt et al., 2009*). *Cushing (1959)* qualitatively described a single peak at higher latitudes and a double peak at lower temperate latitudes, which is consistent with the PFT observations summarized here, as well as the ecosystem model of *Platt et al. (2009)*. Chl time series from SeaWiFS in the North Atlantic analyzed in *Platt et al. (2009)* are also generally consistent with these observations, as is the analysis of *Cabré et al. (2016)*. In general the zonal bands of single vs. double peak run slightly SW to NE in the Northern Hemisphere, which is most pronounced in the Pacific and is apparent in both the analyses here (Fig. 5A) and the maps of *Sapiano et al. (2012)*. Importantly, caution should be employed when interpreting results from areas with low seasonal variance (Sect. 3.7, Figs. 2A and S13A). More details on the secondary blooms in the PFT data sets, including phase difference with the respective primary bloom and fractional prominence analysis, are provided in Supplement Part 7.

The CMIP5 models exhibit a very different pattern of single (Fig. 5C) vs. double peaks (Fig. 5D), as compared to the PFT and Chl satellite data. Double peaks are predominantly found only along the Equator and in the Arabian Sea and the Bay of Bengal. In these areas the models are in agreement with the satellite data (cf. Figs. 5A and 5B), although the data are noisier. Models fail to reproduce the secondary peak occurring at mid-latitudes around $40\text{--}45^\circ$ in satellite data.

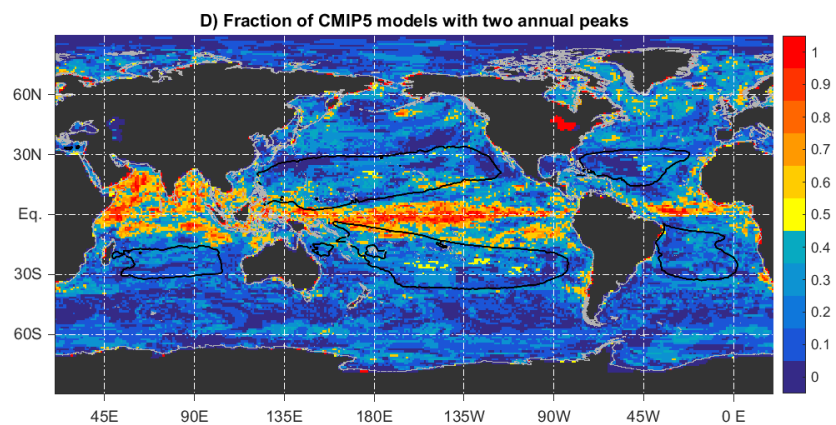
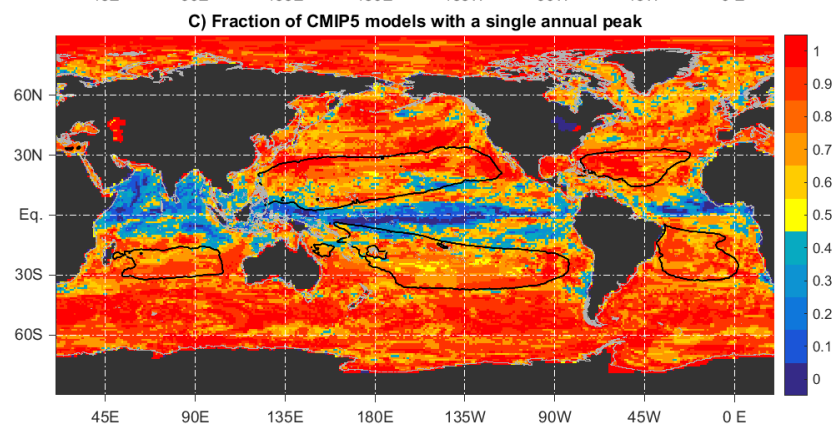
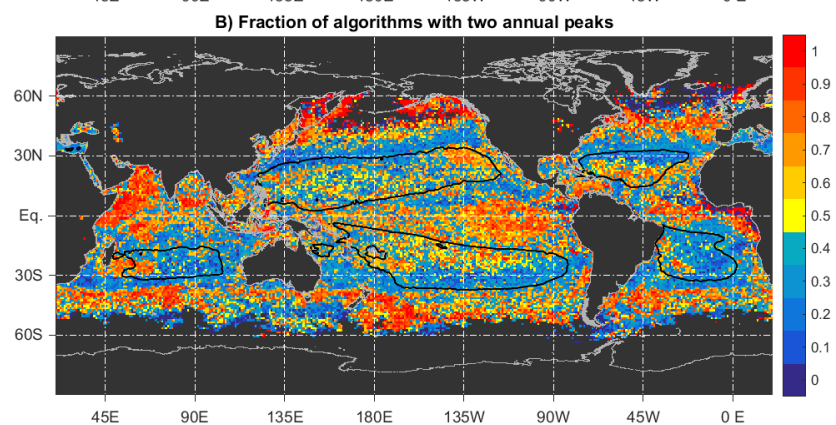
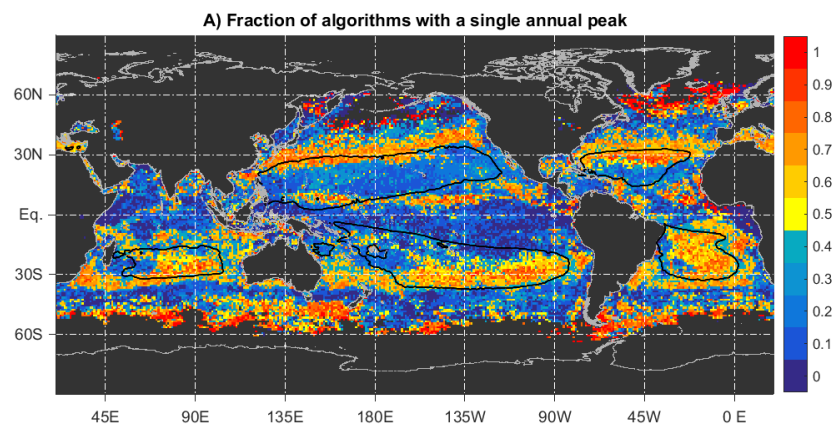


Figure 5. Fraction of PFT algorithms exhibiting (A) a single annual peak or (B) two peaks in one annual cycle. Fraction of CMIP5 algorithms that exhibit (C) a single annual peak and (D) two peaks in one annual cycle. The fraction is calculated from all algorithms (or models, respectively) that have valid phenology metrics calculated at each pixel (Supplement Fig. S6). The isoline of climatological Chl = 0.08 mg m⁻³ (black solid contour) is shown on all panels.

3.6 North Atlantic Regionally Binned Analysis

The analyses of seasonal variance (Fig. 2) and that of number of peaks (Fig. 5) suggest the presence of the following North Atlantic biomes with significant seasonal amplitude: (a) a *subpolar* regime north of about 50°N, with a single (June to August) light-limited biomass peak; and (b) a *transitional*, subpolar-subtropical regime between 35°-50°N with two annual peaks, resulting from an alternation of light and nutrient limited conditions (e.g. *Evans and Parslow (1985)*, and (c) a seasonally varying *Northern subtropics* regime centered around 30°N, characterized by a single annual peak in winter or early spring and high seasonal variability. Monthly averages from regionally-binned satellite PFT time series for two Longhurst marine biogeographic provinces, the North Atlantic Drift Region (NADR) and the North Atlantic Subtropical gyre – West (NASW) (Supplement Fig. S9), are exhibited in Fig. 6A and 6B, respectively. The NADR province straddles the subpolar and the transitional regimes (representing mostly the transitional biome), whereas NASW represents the Northern subtropics regime. The corresponding month of maxima are illustrated for the same Longhurst provinces for the 10 PFT algorithms and Chl (Fig. 7A-B) and for the 7 CMIP5 models (Fig. 7C-D). The time series in both regions (Fig. 6) illustrate that the PFT algorithms exhibit different absolute values

533 and amplitudes/ranges of their variables (Table 1), even if they are mostly referred to as
534 large/micro phytoplankton. This is expected since the algorithms have differing theoretical bases.
535 The spectra of the corresponding complete 5-year time series (Supplement Fig. S10) have their
536 strongest peak at $f = 1 \text{ yr}^{-1}$, indicating that the annual seasonal cycle is a first order source of
537 variability. 2nd and 3rd harmonics often represent additional notable peaks.

538

539

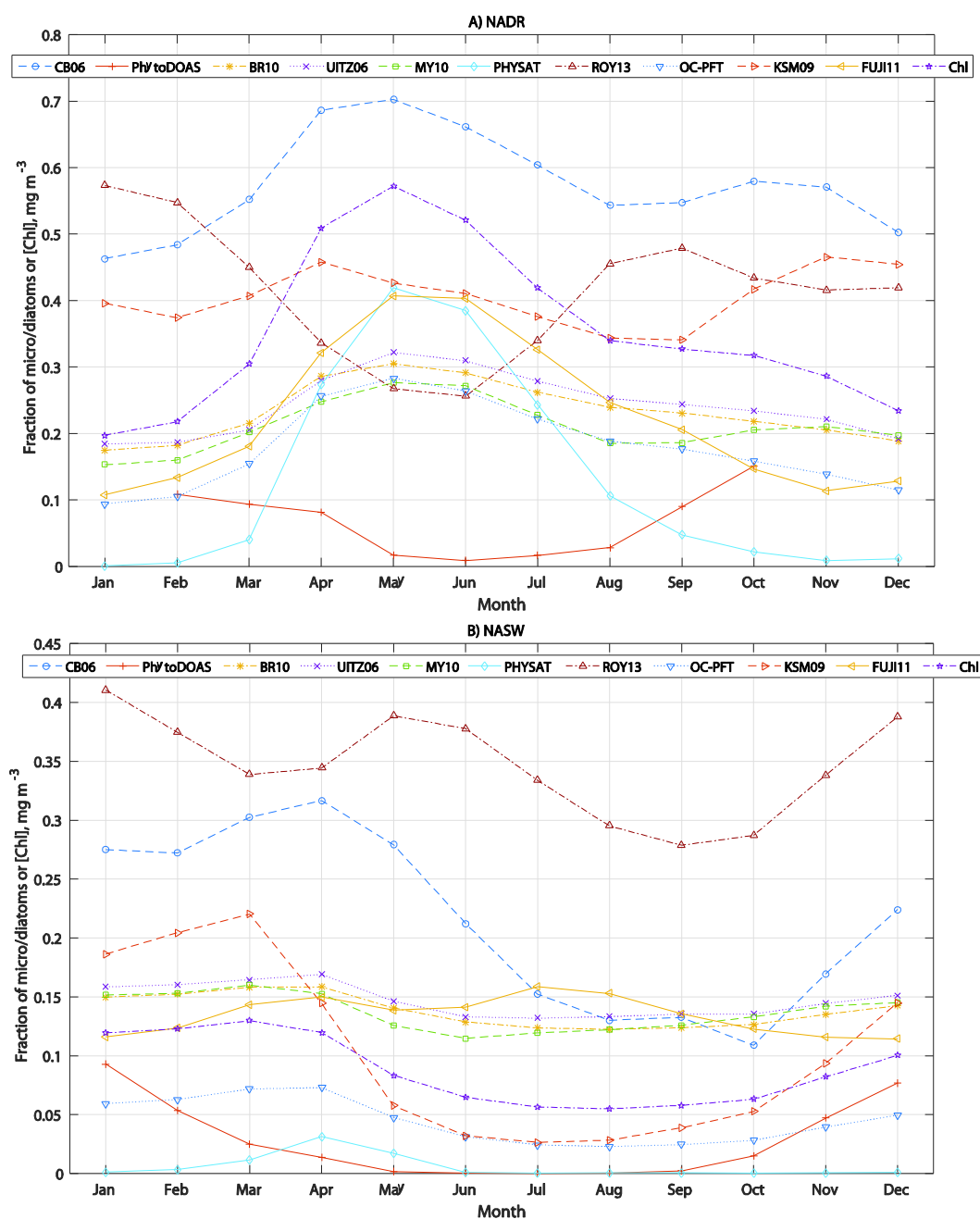


Figure 6. Time series plots of the monthly climatologies of the PFT algorithm variables and OC4v6 Chl (units given in Table 1) and for two example *Longhurst (1998)* biogeographic provinces as follows: A) North Atlantic Drift Region (NADR); B) The Western North Atlantic Subtropical Gyral Province (NASW). Y-axis scales are different

between the two panels. See Sect. 3.6 and Supplement Parts 4 and 8 for more information and Supplement Fig. S9 for a map of the provinces.

Most algorithms in Longhurst's NADR province (Supplement Fig. S9) agree that there is a well pronounced annual maximum in May (Figs. 6A, 7A); however ROY13 and PhytoDOAS indicate minima then instead, exhibiting very different phasing of the seasonal cycle. Also KSM09 exhibits relatively small range of the seasonal cycle there compared to other algorithms, as well as a double peak in April and November (Fig. 7A). The reasons for the lack of complete agreement among the PFT algorithms regarding the month of maximum could not be resolved by this study and require further investigation; this suggests that more *in situ* validation and algorithm development data is required. Three of the seven CMIP5 models also place the maximum in May (Fig. 7C), but for some it is in April or June, and December for the GISS-E2-H-CC model, resulting in a higher variance of month of maximum for the models as compared to the satellite data for NADR (cf. length of black arrows in Figs. 7A and 7C). Some inter-model and data-model differences could be due to the relatively small size of the Longhurst provinces with respect to the coarse model resolutions and the fact that some models may place the equivalent biome in a different location due to different model physics. The NADR province straddles regions where most satellite algorithms indicate double peaks (Figs. 5A and 5B); while CMIP5 models show single annual peaks (Fig. 5C). The fall peak in satellite data is generally weaker than the spring one (Fig. 6A); CB06, KSM09 and MY10 exhibit more noticeable fall blooms. Note that these fall peaks may or may not be detected by the DFT analysis here based on prominence criteria (Sect. 2.3). The presence of double peaks is reflected in the spectra having a pronounced peak at $f = 2 \text{ yr}^{-1}$ (Supplement Fig. S10) that is almost as high as the primary peak at

$f = 1 \text{ yr}^{-1}$. The NADR spectra indeed exhibit higher overall variance (more power) than the NASW spectra, and also more even distribution of power between the annual and semi-annual peaks, indicating the NASW area is characterized by a single annual peak, and NADR has a fall secondary peak. The strength of this peak in relation to the primary one may depend on the unit (fractional vs. absolute) used in the time series analysis (Sect 3.7 and Supplement Part 7). The high variance of NADR is expected, as this area is known for its spectacular North Atlantic blooms in the spring (e.g. *Siegel et al., 2002; Behrenfeld, 2010*). Since NADR straddles two different regimes with respect to single vs. double peaks (cf. Figs. 5A and 5B and Supplement Fig. S9), this analysis illustrates the limitations of a regionally binned approach using classically defined biogeographic provinces.

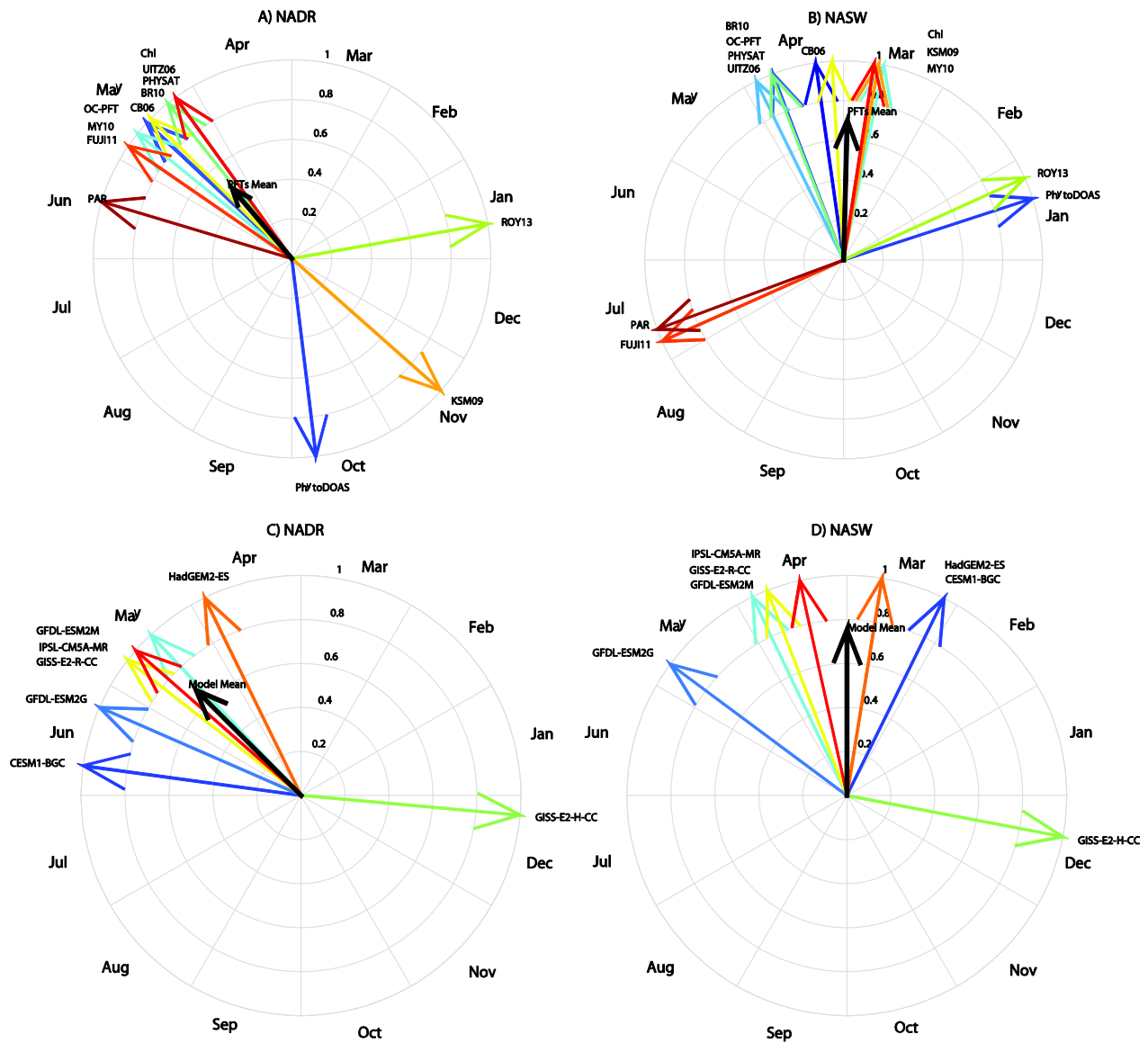


Figure 7. Months of maxima of the regionally binned PFT algorithm variables (Table 1) (top panels – A and B) and CMIP5 models' diatom carbon biomass (bottom panels – C and D). The following Longhurst (1998) provinces are displayed as examples: A and C) North Atlantic Drift Region (NADR); B and D) The Western North Atlantic Subtropical Gyral Province (NASW). See Supplement Fig. S9 for a map of the provinces. The black arrow's direction indicates the PFT algorithms' or CMIP5 models' ensemble mean month

of maximum (circular), and its length indicates one minus the circular variance in month of maxima among the algorithms (an arrow length of one indicates zero variance, and an arrow length of zero – maximal variance of one). The beginning of each month is marked by the abbreviated month name on the polar plots. See Sect. 2.3 and Supplement Part 4 for methodology details. Supplement Table S1 lists the CMIP5 models used with their acronyms as they appear here.

Further to the south, the NASW province straddles the $\text{Chl} = 0.08 \text{ mg m}^{-3}$ climatological isoline used here to delineate the gyre (Supplement Fig. S9) and is a typical northern-subtropical region. It is mostly characterized by a single peak in most algorithms and exhibits some of the cleanest seasonal cycles globally (Figs. 2A and 5A). It is considerably more oligotrophic and hence has lower fraction of microplankton (and total Chl concentration) than NADR, year-round (cf. y-axis scales of Fig. 6A vs. 6B). While some algorithms indicate a strong winter-spring peak (KSM09, CB06), the annual range of others is a lot smaller and some even exhibit double peaks (ROY13, FUJI11). Seven algorithms agree on a maximum for the primary bloom in March or April (Fig. 7B), but ROY13 and PhytoDOAS indicate a January peak, and FUJI11 indicates a July peak for the primary bloom. Similarly, most CMIP5 models agree on a March or April bloom peak (Fig. 7D). As stated above, the DFT spectra of the satellite data (Supplement Fig. S10B) indicate that overall variance is lower than NADR, and the first harmonic contains proportionately more power than the second harmonic, indicating a single annual peak. A comparative analysis for time-series at the Bermuda Atlantic Time Series (BATS) station, representative of the NASW province and the Northern subtropics regime, is provided in Supplement Part 8 (Supplement Fig.

S11), including discussion of some complementary *in situ* biogeochemical data. Supplement Part 8 also provides details on mechanisms in the regions of interest discussed here.

In conclusion, regional binning of the satellite data sets reduces noise, and allows for inspection of a limited number of actual time series from various regions. It also allows for analysis in regions or algorithms that suffer from data sparsity (particularly true of the PHYSAT algorithm) that precludes the DFT analysis on a per-pixel basis. On the other hand, if the chosen regions are spatially heterogeneous, results may be misleading or meaningless, and regions may have an arbitrary definition not necessarily relevant to the phenology at question here. Future monitoring of biology and biogeochemistry at multiple specific representative locations such as BATS is critically important for validation and inter-comparison of satellite algorithms.

3.7 Sources of Uncertainty

There are multiple sources of uncertainty that can affect the DFT phenology analysis presented, both related to intrinsic ecosystem characteristics and features and limitations of the DFT technique. Here we discuss some important considerations and we provide more details in Supplement Part 9. The advantages of the DFT method include relative insensitivity to noise and the ability to isolate the variance in the frequencies of interest for phenological studies (*deBeurs and Henebry, 2010*). In addition, multi-year time series are summarized with a single metric from the modeled signal and the DFT approach eliminates the need to consider a "sliding season" to ensure the annual cycle is properly described everywhere (*Racault et al., 2012*). The DFT is a statistical method to fit data to a sum of sines and cosines of fixed frequencies and varying phases. As such, one disadvantage is that representation of seasonal cycles that have non-

sinusoidal waveforms requires artificial placement of variance in high frequencies where most noise resides. For example, *Wilson and Qiu, 2008* observe sharp spikes in seasonal Chl blooms in some areas of the oligotrophic gyres. If higher frequencies are present in the data, significant aliasing can occur, confounding the analysis by placing spurious energy in lower frequencies. The aliasing problem can be addressed by analyzing weekly or daily data to test for the significance of higher order harmonics in the data. That said, Fourier analysis is among the best available techniques to analyze cyclical phenomena and partition variance in frequency components of interest, so it is strongly suited for analyzing seasonal cycles.

The correct retrieval of the phenological parameters of interest using the DFT technique was verified in several ways: by examining an example time series (Supplement Fig. S1); by validation against direct maxima finding via peak analysis of the time series of the monthly climatologies of the respective PFT satellite data, and relating the results to percent seasonal variance (Supplement Fig. S12); and by using the SeaWiFS PAR data set (as discussed in Supplement Sect. S1). The overall assessment is that the DFT technique correctly identifies phenological parameters if percent variance explained by the seasonal harmonics is 30% or more. Phase derived by the DFT techniques in frequency bands where power density is low can be stochastic and meaningless; therefore month of maxima determinations can be unreliable or random and should be treated with caution in areas of low percent seasonal variance (Fig. 2; Supplement Fig. S13). Supplement Fig. S12A illustrates the fraction of pixels for each algorithm that exhibit a given percent seasonal variance. Comparison of direct maxima finding vs. the DFT determinations of the month of maxima indicate that in places where percent seasonal variance drops below 30%, differences between the two methods can be > 2 months for over 10% of the

pixels for most algorithms (Supplement Fig. S12B). It was therefore determined that the DFT technique results should be interpreted with caution or not used in analyses in places where percent seasonal variance drops below 30%. A map of the number of algorithms exhibiting percent seasonal variance $< 30\%$ is shown in Supplement Fig. S13A, and the analogous map for the CMIP5 models is shown in Supplement Fig. S13B. It is worth noting that analyzing phenology in places with low percent seasonality becomes intrinsically meaningless, regardless of what technique is used. Additionally, ensemble metrics for the PFT algorithms should be interpreted with caution where few algorithms contribute to the mean (Supplement Fig. S6). Finally, results should also be interpreted with caution at high latitudes where data can be sparse, especially in the respective winter months. This is discussed in Supplement Sect. S1.

It is important to note that it is not exactly equivalent to analyze absolute biomass indicators (Chl, carbon) and fractions (as done here for most PFT algorithms), e.g. there are indications that the secondary bloom is more dominated by larger sizes than the primary peak (*Cabré et al., 2016*) (see also Supplement Part 5 and Part 7). For example, a secondary peak can look about as high as the primary one in terms of percent, but can be much weaker in terms of Chl. Mathematically, percent microplankton can increase without an accompanying increase (or even with a decrease) in total or microplankton absolute biomass or Chl. However, such situations are likely to be atypical on a global scale according to modern ecosystem understanding. Absolute carbon biomass of diatoms was analyzed for the CMIP5 models, which is not exactly equivalent to the algorithm variables. Further analysis should focus on comparing phenology calculated from fractions vs. absolute biomass indicators. Finally, physiological adaptation affects the Chl to carbon ratio and can decouple Chl and carbon variability, especially in lower-latitude oceans (e.g.

Behrenfelred et al., 2005; Siegel et al., 2013); it is best to express size-fractionated or PFT-specific biomass in terms of carbon units, as done by *Kostadinov et al., 2016* who re-cast the PFTs in terms of carbon using the KSM09 particle size distribution algorithm and allometric relationships (*Menden-Deuer and Lessard, 2000*). In spite of the above considerations, we stress that the primary purpose of this work is PFT algorithm inter-comparison, and all PFT algorithms are analyzed in an equivalent way here.

The phenological parameters described here were derived only from the large phytoplankton/microplankton variable (Table 1) or for diatoms in the case of the CMIP5 models and the PhytoDOAS and PHYSAT algorithms. While in much of the temperate, high latitude and upwelling regions it is indeed larger phytoplankton that dominate the bloom, in the more oligotrophic subtropics and tropics nanoplankton can dominate the seasonal maximum. Also, diatoms can be found within the nanoplankton fraction or the microplankton can be dominated by other large phytoplankton (e.g. dinoflagellates). Various PFT algorithms provide nanoplankton and other PFT variables that need to be analyzed and compared in future work. The analysis presented here is global; however, some algorithms were developed/parameterized with data sets of specific limited geographic coverage. Those would not be expected to necessarily perform well outside of their area of development. Notably, FUJI11 was developed for the Arctic. Not surprisingly, this algorithm differs more substantially from other algorithms in the tropics and subtropics. The Southern Ocean presents atypical bio-optical characteristics (e.g. *Uitz et al., 2006*) some implications of which are discussed in Supplement Part 9.

4. Concluding Remarks

We used the Discrete Fourier Transform (DFT) to derive and inter-compare phenological parameters for the 2003-2007 period among 1) the fraction of Chl corresponding to microplankton (or a closely related variable) from 10 satellite ocean color algorithms, 2) satellite determinations of chlorophyll concentration, and 3) diatom biomass from 7 CMIP5 climate models. The phenological parameters derived were amplitude, month of maximum, percent variance explained by the seasonal cycle, bloom duration, and secondary bloom characteristics. Results indicate that PFT algorithms agree only to first order globally. Enough qualitative and quantitative differences between the algorithms are detected (e.g. Fig. 6) to make a further comprehensive, global validation exercise a high priority. While validation is outside the scope of this work, a separate working group has been formed within the PFT Inter-comparison Project to perform a comprehensive validation exercise (*Bracher et al., 2015*). Validation itself is challenging (*Brewin et al., 2011*), as *in situ* HPLC-derived PFTs have their own limitations and do not necessarily correspond to the way non-HPLC-based algorithms define their variables. To allow for these subtle differences in the variables retrieved, it may be best to test all algorithms against a comprehensive *in situ* data set incorporating co-located radiometric, bio-optical (pigment concentration, IOPs) and derived biological quantities (phytoplankton Chl, size structure, etc.) in order to resolve whether differences are due to algorithm uncertainties or to actual biogeophysical differences between, for example, the timing of maximum diatom biomass vs. maximum of large cells biomass (*Bracher et al. 2015*).

Comparison of phenological parameters in CMIP5 model output (diatom carbon biomass) to those of satellite data suggests that the month of maximum is fairly well represented in models

(albeit with a systematic, latitude-dependent bias), while other phenological characteristics show a number of important biases in CMIP5 models: 1) more pronounced seasonal variability in the models, e.g. a smoother latitudinal progression and less local spatial variability in phenological indices such as month of maximum and bloom duration, 2) while seasonality in the satellite data is cleanest (i.e. percent variance explained is highest) along a zonal band at 30° latitude in both, this feature is not reflected in the CMIP5 models. and 3) models exhibit a single annual biomass peak over most of the ocean, except for the Equatorial band, whereas secondary blooms tend to occur in zonal bands in temperate latitudes in the satellite data but are not well captured in the climate models. These biases are probably due to over-simplification of processes in models and a lower response to interannual variability than in reality as also discussed in *Cabré et al. (2016)*. Additionally, the coarse 1° resolution does not allow a proper representation of coastal processes and some frontal dynamics in models. We note that for many modelers, the PFT products derived from satellite algorithms are considered as observations (not algorithm products per se) and used as a reference for validating model outputs, (e.g. *LeQuéré et al. (2005)*; *Bopp et al. (2005)*; *Stock et al. (2014)*). To improve the utility of satellite algorithms for this purpose, better uncertainty characterization should be considered high priority.

A fruitful way forward would be to use algorithms of different theoretical bases together to increase the degrees of freedom and solve for more variables. Advent of hyperspectral sensors in the near future is expected to improve our ability to discern small spectral differences arising from the different PFTs, as indicated, for example, by the development of the PhytoDOAS algorithm. Therefore, future direction of efforts towards development of hyperspectral algorithms is desirable, keeping in mind that there is a fundamental limit on the additional

degrees of freedom available (*Lee et al., 2007*). While chlorophyll is certainly a useful variable, it is carbon biomass in the living phytoplankton that is the variable of most direct relevance to carbon cycle and biogeochemical studies; it is also the unit of PFT accounting in climate models (Table S1). The carbon-based algorithm of *Kostadinov et al. (2016)* could be used in conjunction with an algorithm partitioning Chl (e.g. BR10) in order to assess physiological status and productivity by size class (*Behrenfeld et al., 2005; Uitz et al., 2010*).

Importantly, PFT algorithms and bio-optical algorithms in general could improve by moving towards analytical approaches based more on first principles rather than empirical relationships, i.e. being mechanistic in nature. Most of the existing PFT algorithms contain a high degree of empiricism. Empirical algorithms rely on statistical relationships derived during a certain environmental state and are thus not predictive in nature. Should the underlying relationship change, the algorithm uncertainties will increase. Mechanistic models should remain more robust under changing environmental conditions of the future, e.g. due to climate change.

Acknowledgments

This work was performed with funding from NASA Ocean Biology and Biogeochemistry Program (grant #NNX13AC92G to Irina Marinov and Tihomir S. Kostadinov). We thank Tilman Dinter and Bernard Gentili for help with the PhytoDOAS and CB06 algorithm processing, respectively, Amane Fujiwara for leading development of the FUJI11 algorithm, and Aurea Ciotti for leading the development of the CB06 algorithm. We thank David Shields (specifically for producing Supplement Fig. S11C and processing BATS data, and for work on the Chl gyre contour lines), Svetlana Milutinović, and Danica Fine for providing help in the completion of this

work. We thank Libe Washburn for FFT processing advice. We thank Jordan Rosenthal for his compass plot labeling script (used here with modifications in Fig. 7). All data processing, analysis and plotting for the phenological analysis in this work was done in MATLAB®. We additionally acknowledge all the satellite algorithm providers (and their funding agencies) for their support and providing their data. The coastlines displayed on maps were extracted with the NOAA/NGDC GEODAS-NG software using the L1 layer of the GSHHG v2.2.3 (*Wessel and Smith, 1996*) coastline data set. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table S1 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Inter-comparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We are grateful to three anonymous reviewers and the editor whose comments and suggestions improved the quality of this manuscript.

References

- Alvain, S., Moulin, C., Dandonneau, Y., & Bréon, F. M. (2005), Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery. *Deep Sea Research Part I: Oceanographic Research Papers*, 52(11), 1989-2004.
- Alvain, S., C. Moulin, Y. Dandonneau, and H. Loisel (2008), Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: A satellite view, *Global Biogeochem. Cycles*, 22, GB3001, doi:10.1029/2007GB003154.

790

791 Behrenfeld, M. J., E. Boss, D. A. Siegel, and D. M. Shea (2005), Carbon-based ocean
 792 productivity and phytoplankton physiology from space, *Global Biogeochem. Cycles*, 19,
 793 GB1006, doi:10.1029/2004GB002299.

794

795 Behrenfeld, M. J. (2010). Abandoning Sverdrup's critical depth hypothesis on phytoplankton
 796 blooms. *Ecology*, 91(4), 977-989.

797

798 Bograd, S. J., D. G. Foley, F. B. Schwing, C. Wilson, R. M. Laurs, J. J. Polovina, E. A. Howell,
 799 and R. E. Brainard (2004), On the seasonal and interannual migrations of the transition zone
 800 chlorophyll front, *Geophys. Res. Lett.*, 31, L17204, doi:10.1029/2004GL020637.

801 Bopp, L., Aumont, O., Cadule, P., Alvain, S., & Gehlen, M. (2005). Response of diatoms
 802 distribution to global warming and potential implications: A global model study. *Geophysical*
 803 *Research Letters*, 32(19), L19606. doi:10.1029/2005GL023653

804 Bracher, A., Vountas, M., Dinter, T., Burrows, J. P., Röttgers, R., and Peeken, I. (2009),
 805 Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on
 806 SCIAMACHY data, *Biogeosciences*, 6, 751-764, doi:10.5194/bg-6-751-2009.

807

808 Bracher N., Hardman-Mountford N., Hirata T., Bernard S., Brewin R., Bricaud A., Brotas V.,
 809 Chase A., Ciotti A., Choi J.-K., Clementson L., Devred E., DiGiacomo P., Dupouy C., Kim W.,
 810 Kostadinov T., Kwiatkowska E., Lavender S., Moisan T., Mouw C., Son S., Sosik H., Uitz J.,
 811 Werdell J. , Zheng G. (2015) Report on IOCCG workshop “Phytoplankton composition from

812 Space: towards a validation strategy for satellite algorithms” 25 - 26 Oct 2014, Portland, ME.,
 813 USA, <http://www.ioccg.org/groups/report-PFTworkshopOct2014.pdf>; as NASA technical
 814 Memorandum 217528_01-22-15 [http://www.ioccg.org/groups/PFT-TM_2015-217528_01-22-](http://www.ioccg.org/groups/PFT-TM_2015-217528_01-22-15.pdf)
 815 [15.pdf](http://www.ioccg.org/groups/PFT-TM_2015-217528_01-22-15.pdf)
 816
 817 Brewin, R. J., Sathyendranath, S., Hirata, T., Lavender, S. J., Barciela, R. M., & Hardman-
 818 Mountford, N. J. (2010), A three-component model of phytoplankton size class for the Atlantic
 819 Ocean, *Ecological Modelling*, 221(11), 1472-1483.
 820
 821 Brewin, R. J., Hardman-Mountford, N. J., Lavender, S. J., Raitsos, D. E., Hirata, T., Uitz, J.,
 822 Devred, E., Bricaud, A., Ciotti, A. and Gentili, B. (2011). An inter-comparison of bio-optical
 823 techniques for detecting dominant phytoplankton size class from satellite remote sensing. *Remote*
 824 *Sensing of Environment*, 115(2), 325-339.
 825
 826 Bricaud, A., A. M. Ciotti, and B. Gentili (2012), Spatial-temporal variations in phytoplankton
 827 size and colored detrital matter absorption at global and regional scales, as derived from twelve
 828 years of SeaWiFS data (1998–2009), *Global Biogeochem. Cycles*, 26, GB1010,
 829 doi:10.1029/2010GB003952.
 830
 831 Cabré, A., I. Marinov and S. Leung (2015), Consistent global responses of marine ecosystems
 832 to future climate change across the IPCC AR5 earth system models, *Clim Dyn* 45: 1253.
 833 doi:10.1007/s00382-014-2374-3.
 834

835 Cabré A., D. Shields, I. Marinov and T.S. Kostadinov (2016), Phenology of size-partitioned
836 phytoplankton carbon-biomass from ocean color remote sensing and CMIP5 models, *Front. Mar.*
837 *Sci.*, 3:39. doi: 10.3389/fmars.2016.00039.

838

839 Campbell, J. W. (1995). The lognormal distribution as a model for bio-optical variability in the
840 sea. *Journal of Geophysical Research: Oceans* (1978–2012), 100(C7), 13237-13254.

841

842 Chisholm, S. W. (1992), Phytoplankton size. In P. G. Falkowski and A. D. Woodhead (Eds.),
843 Primary Productivity and Biogeochemical Cycles in the Sea. New York, Plenum Press, p. 213-
844 237.

845

846 Ciotti, A. M., & Bricaud, A. (2006), Retrievals of a size parameter for phytoplankton and spectral
847 light absorption by colored detrital matter from water-leaving radiances at SeaWiFS channels in a
848 continental shelf region off Brazil, *Limnology and Oceanography: Methods*, 4, 237-253.

849

850 Cushing, D.H. (1959), The seasonal variation in oceanic production as a problem in population
851 dynamics. *Journal du Conseil*, 24 (3), 455–464.

852

853 Eppley, R. W., & Peterson, B. J. (1979), Particulate organic matter flux and planktonic new
854 production in the deep ocean. *Nature*, 282, 677

855

856 Evans, G.T. and J. S. Parslow (1985), A Model of Annual Plankton Cycles, *Biological*
857 *Oceanography*, 3:3, 327-347.

858

859 Falkowski, P.G., R.T. Barber, V. Smetacek (1998), Biogeochemical controls and feedbacks on
860 ocean primary production, *Science*, 281, 200–206.

861

862 Field, CB, Behrenfeld MJ, Randerson JT and Falkowski P (1998), Primary production of the
863 biosphere: Integrating terrestrial and oceanic components, *Science*, 281: 237–240.

864

865 Fujiwara, A., Hirawake, T., Suzuki, K., and Saitoh, S.-I. (2011), Remote sensing of size structure
866 of phytoplankton communities using optical properties of the Chukchi and Bering Sea shelf
867 region, *Biogeosciences*, 8, 3567-3580, doi:10.5194/bg-8-3567-2011.

868

869 Glover, D. M., J. S. Wroblewski, and C. R. McClain (1994), Dynamics of the transition zone in
870 coastal zone color scanner-sensed ocean color in the North Pacific during oceanographic spring,
871 *J. Geophys. Res.*, 99, 7501– 7511.

872

873 Habib, E. A. (2012). Geometric mean for negative and zero values. *Int. J. Res. Rev. Appl. Sci*, 11,
874 419-432.

875

876 Hirata, T., Hardman-Mountford, N. J., Brewin, R. J. W., Aiken, J., Barlow, R., Suzuki, K., Isada,
877 T., Howell, E., Hashioka, T., Noguchi-Aita, M., and Yamanaka, Y., (2011), Synoptic
878 relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton
879 functional types, *Biogeosciences*, 8, 311-327, doi:10.5194/bg-8-311-2011.

880

881 Hirata, T., N. Hardman-Mountford and R. J. W. Brewin, (2012), Comparing satellite-based
 882 phytoplankton classification methods, *Eos Trans. AGU*, 93(6), 59.
 883

884 Hirata, T. (2015), Satellite Phytoplankton Functional Type Algorithm Inter-comparison Project,
 885 <http://pft.ees.hokudai.ac.jp/satellite/index.shtml>, last access: Oct. 16, 2015.
 886

887 Hood, R. R., E.A. Laws, R.A. Armstrong, N.R. Bates, C.W. Brown, C.A. Carlson, F. Chai, S.C.
 888 Doney, P.G. Falkowski, R.A. Feely, M.A.M. Friedrichs, M.R. Landry, J.K. Moore, D.M. Nelson,
 889 T.L. Richardson, B. Salihoglu, M. Schartau, D.A. Toole, J.D. Wiggert (2006), Pelagic functional
 890 group modeling: Progress, challenges and prospects, *Deep-Sea Research II*, 53, 459–512.
 891

892 IOCCG (2014). Phytoplankton Functional Types from Space. Sathyendranath, S. (ed.), Reports
 893 of the International Ocean-Colour Coordinating Group, No. 15, IOCCG, Dartmouth, Canada.
 894

895 IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I
 896 to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F.,
 897 D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M.
 898 Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY,
 899 USA, 1535 pp.
 900

901 Koeller, P., Fuentes-Yaco, C., Platt, T., Sathyendranath, S., Richards, A., Ouellet, P., ... &
 902 Aschan, M. (2009). Basin-scale coherence in phenology of shrimps and phytoplankton in the
 903 North Atlantic Ocean. *Science*, 324(5928), 791-793.

904

905 Kostadinov, T. S., D. A. Siegel, and S. Maritorena (2009), Retrieval of the particle size
 906 distribution from satellite ocean color observations, *J. Geophys. Res.*,
 907 doi:10.1029/2009JC005303.

908

909 Kostadinov, T. S., D. A. Siegel, and S. Maritorena (2010), Global variability of phytoplankton
 910 functional types from space: assessment via the particle size distribution, *Biogeosciences*, 7(10),
 911 3239-3257.

912

913 Kostadinov, T. S., Milutinović, S., Marinov, I., and Cabré, A. (2016), Carbon-based
 914 phytoplankton size classes retrieved via ocean color estimates of the particle size distribution,
 915 *Ocean Sci.*, 12, 561-575, doi:10.5194/os-12-561-2016.

916

917 Le Quéré, C., S.P. Harrison, I. C. Prentice, E.T. Buitenhuis, O. Aumont, L. Bopp, H. Claustre, L.
 918 Cotrim Da Cunha, R. Geider, X. Giraud, C. Klaas, K.E. Kohfeld, L. Legendre, M. Manizza, T.
 919 Platt, R.B. Rivkin, S. Sathyendranath, J. Uitz, A.J. Watson, D. Wolf-Gladrow (2005), Ecosystem
 920 dynamics based on plankton functional types for global ocean biogeochemistry models. *Global*
 921 *Change Biology 11*, 2016–2040.

922

923 Lee, ZhongPing , Kendall Carder, Robert Arnone and MingXia He (2007), Determination of
 924 Primary Spectral Bands for Remote Sensing of Aquatic Environments. *Sensors* 2007, 7, 3428-
 925 3441.

926

927 Levy, M., Lehahn, Y., Andre, J. M., Memery, L., Loisel, H., and Heifetz, E. (2005), Production
 928 regimes in the northeast Atlantic: A study based on Sea-viewing Wide Field-of-view Sensor
 929 (SeaWiFS) chlorophyll and ocean general circulation model mixed layer depth. *J. Geophys. Res.*
 930 110. doi:10.1029/2004jc002771.
 931
 932 Loisel, H., J.-M. Nicolas, A. Sciandra, D. Stramski, and A. Poteau (2006), Spectral dependency
 933 of optical backscattering by marine particles from satellite remote sensing of the global ocean, *J.*
 934 *Geophys. Res.*, 111, C09024, doi:10.1029/2005JC003367.
 935
 936 Longhurst, A.R. (1998). *Ecological Geography of the Sea*. Academic Press, San Diego. 397p.
 937
 938 McClain, C.R. (2009), A decade of satellite ocean color observations, *Annual Review of Marine*
 939 *Science*, 1, 19-42.
 940
 941 Marañón, E. (2015), Cell Size as a Key Determinant of Phytoplankton Metabolism and
 942 Community Structure, *Annu. Rev. Mar. Sci.*, 7(1), doi: 10.1146/annurev-marine-010814-015955.
 943
 944 Menden-Deuer S. and Lessard EJ (2000), Carbon to volume relationships for dinoflagellates,
 945 diatoms, and other protist plankton, *Limnol. Oceanogr.*, 45: 569–579.
 946
 947 Mouw, C. B., and J. A. Yoder (2010), Optical determination of phytoplankton size composition
 948 from global SeaWiFS imagery, *J. Geophys. Res.*, 115, C12018, doi:10.1029/2010JC006337.
 949

950 NASA Goddard Space Flight Center, Ocean Biology Distributed Active Archive Center; (2010):
 951 Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Ocean Color Data, NASA OB.DAAC,
 952 Greenbelt, MD, USA. Reprocessing R2010.0. Maintained by NASA Ocean Biology Distributed
 953 Active Archive Center (OB.DAAC), Goddard Space Flight Center, Greenbelt MD.
 954
 955 NCEI: National Centers for Environmental Information (NOAA) (2015), Registration of
 956 Structured Square-Cell Grids, URL:
 957 <https://www.ngdc.noaa.gov/mgg/global/gridregistration.html>, last accessed: Dec. 8, 2015.
 958
 959 O'Reilly, J.E., Maritorena, S., Mitchell, B.G., Siegel, D.A., Carder, K.L., Garver, S.A., Kahru,
 960 M., McClain, C. (1998), Ocean chlorophyll algorithms for SeaWiFS, *J. Geophys. Res.* 103(C11):
 961 24,937–24,953.
 962
 963 O'Reilly, J.E., Maritorena, S., Siegel, D.A., O'Brien, M.C., Toole, D., Mitchell, B.G., et al.
 964 (2000). Ocean color chlorophyll-a algorithms for SeaWiFS, OC2, and OC4: version 4, SeaWiFS
 965 postlaunch calibration and validation analyses, Part 3, NASA/TM 206892, 11, 9–23.
 966
 967 Platt, T., Fuentes-Yaco, C., & Frank, K. T. (2003). Marine ecology: spring algal bloom and larval
 968 fish survival. *Nature*, 423(6938), 398-399.
 969 Platt, T., White III, G. N., Zhai, L., Sathyendranath, S., & Roy, S. (2009). The phenology of
 970 phytoplankton blooms: Ecosystem indicators from remote sensing. *Ecological Modelling*,
 971 220(21), 3057-3069.
 972

973 Racault, M. F., Le Quéré, C., Buitenhuis, E., Sathyendranath, S., & Platt, T. (2012).
 974 Phytoplankton phenology in the global ocean. *Ecological Indicators*, 14(1), 152-163.
 975
 976 Roy, S., Sathyendranath, S., & Platt, T. (2011), Retrieval of phytoplankton size from bio-optical
 977 measurements: theory and applications, *Journal of The Royal Society Interface*, 8(58), 650-660.
 978
 979 Roy, S., Sathyendranath, S., Bouman, H., & Platt, T. (2013), The global distribution of
 980 phytoplankton size spectrum and size classes from their light-absorption spectra derived from
 981 satellite data, *Remote Sensing of Environment*, 139, 185-197.
 982
 983 Sadeghi A., Dinter T., Vountas M., Taylor B., Peeken I., Altenburg Soppa M., Bracher A. (2012),
 984 Improvements to the PhytoDOAS method for identification of coccolithophores using hyper-
 985 spectral satellite data. *Ocean Sciences* 8: 1055-1070.
 986
 987 Sapiano, M. R. P., C. W. Brown, S. Schollaert Uz, and M. Vargas (2012), Establishing a global
 988 climatology of marine phytoplankton phenological characteristics, *J. Geophys. Res.*, 117,
 989 C08026, doi:10.1029/2012JC007958.
 990
 991 Sieburth, J. M., V. Smetacek, and J. Lenz (1978), Pelagic ecosystem structure: heterotrophic
 992 compartments of the plankton and their relationship to plankton size fractions, *Limnol.*
 993 *Oceanogr.*, 23, 1256-1263.
 994

995 Siegel, D. A., Doney, S. C., & Yoder, J. A. (2002). The North Atlantic spring phytoplankton
 996 bloom and Sverdrup's critical depth hypothesis. *science*, 296(5568), 730-733.
 997
 998 Siegel, D. A., M. J. Behrenfeld, S. Maritorena, C. R. McClain, D. Antoine, S. W. Bailey, P. S.
 999 Bontempi, E. S. Boss, H. M. Dierssen, S. C. Doney, R. E. Eplee Jr, R. H. Evans, G. C. Feldman,
 1000 E. Fields, B. A. Franz, N. A. Kuring, C. Mengelt, N. B. Nelson, F. S. Patt, W. D. Robinson, J. L.
 1001 Sarmiento, C. M. Swan, P. J. Werdell, T. K. Westberry, J. G. Wilding, and J. A. Yoder (2013),
 1002 Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission, *Remote*
 1003 *Sensing of Environment*, 135(0), 77-91.
 1004
 1005 Siegel, D. A., K. O. Buesseler, S. C. Doney, S. F. Sailley, M. J. Behrenfeld, and P. W. Boyd
 1006 (2014), Global assessment of ocean carbon export by combining satellite observations and food-
 1007 web models, *Global Biogeochem. Cycles*, 28, 181–196, doi:10.1002/2013GB004743.
 1008
 1009 Stock, Charles A., John P. Dunne, Jasmin G. John (2014), Global-scale carbon and energy flows
 1010 through the marine planktonic food web: An analysis with a coupled physical–biological model,
 1011 *Progress in Oceanography*, Volume 120, Pages 1-28, doi: 10.1016/j.pocean.2013.07.001.
 1012
 1013 Uitz, J., H. Claustre, A. Morel, and S. B. Hooker (2006), Vertical distribution of phytoplankton
 1014 communities in open ocean: An assessment based on surface chlorophyll, *J. Geophys. Res.*, 111,
 1015 C08005, doi:10.1029/2005JC003207.
 1016

1017 Uitz, J., H. Claustre, B. Gentili, and D. Stramski (2010), Phytoplankton class-specific primary
 1018 production in the world's oceans: Seasonal and interannual variability from satellite observations,
 1019 *Global Biogeochem. Cycles*, 24, GB3016, doi:10.1029/2009GB003680.
 1020
 1021 Vidussi, F., H. Claustre, B. B. Manca, A. Luchetta, and J. C. Marty (2001), Phytoplankton
 1022 pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea
 1023 during winter, *J. Geophys. Res.*, 106(C9), 19,939–19,956.
 1024
 1025 Wessel, P., and W. H. F. Smith, A (1996), Global Self-consistent, Hierarchical, High-resolution
 1026 Shoreline Database, *J. Geophys. Res.*, 101, #B4, pp. 8741-8743.
 1027
 1028 Wilson, C., & Qiu, X. (2008). Global distribution of summer chlorophyll blooms in the
 1029 oligotrophic gyres. *Progress in Oceanography*, 78(2), 107-134.

Supplement to Kostadinov et al. " Inter-Comparison of Phytoplankton Functional Type Phenology Metrics Derived from Ocean Color Algorithms and Earth System Models "

Part 1. Details of the DFT Analysis

The per-pixel time series of 1-degree PFT data (Sect. 2.2 and 2.3) is a 60×1 vector \mathbf{x} , consisting of monthly sample points, x_t ; $t = 1, 2, 3, \dots, N$, $N = 60$. For the purposes of Fourier analysis, we consider one year a unit of time, thus the sampling frequency is $f_s = 12$ times per year. The complex-valued discrete Fourier transform (DFT) of \mathbf{x} , denoted \mathbf{y} , was obtained using the Fast Fourier Transform (FFT) algorithm as follows (Heinzel et al., 2002):

$$y_m = \sum_{t=0}^{N-1} x_t e^{\frac{-2\pi i m t}{N}}, m = 0, 1, 2, \dots, N-1 \quad (S1)$$

No windows were applied (i.e. a top-hat window was implicitly applied). The two-sided DFT transform vector \mathbf{y} is composed of the y_m elements. The frequency resolution of the FFT is thus $f_s/N = 0.2 \text{ yr}^{-1}$, and the frequencies to which the elements of \mathbf{y} correspond are $f = 0, 0.2, 0.4, 0.6, \dots \text{ yr}^{-1}$ up to the highest resolvable frequency – the Nyquist frequency $f_{\text{Nyquist}} = f_s/2 = 6 \text{ yr}^{-1}$. The first element of the \mathbf{y} vector is real and equal to the arithmetic mean of the data set \mathbf{x} ; thus it is equal to zero because the mean was subtracted before Eq. S1 was applied. Because of a property of the DFT of real valued input, the first half of vector \mathbf{y} contains the same information as the second half and they are complex conjugates of each other. For example if the 2nd element is $a + bi$, then the last element is $a - bi$. Because of this conjugate symmetry, only the first half of the vector \mathbf{y} has to be considered. The power in each frequency is spread to both complex conjugates, so we need to multiply the square of the modulus of each element of the first half of \mathbf{y} by 2, with the important exceptions of the 1st element (the mean), and in the case of even N – the $(N/2 + 1)$ -th element corresponding to the Nyquist frequency (Heinzel et al., 2002).

The DFT vector is thus normalized as follows, in order to obtain the power spectrum density, P_{xx} (Heinzel et al., 2002) (symbolology as above):

$$P_{xx}(f_m) = \left(\frac{k |y_m|^2}{N f_s} \right); m = 0, 1, 2, \dots, \frac{N}{2} \quad (S2)$$

where $k = 1$ for $m = 0$ (the signal mean, 0th frequency) and $m = N/2$ (the Nyquist frequency), and $k = 2$ for all other m values. P_{xx} has the units of squared data units per cycle per year, that is, power per unit frequency.

The power spectrum, PS , can be obtained by multiplying Eq. S2 by the frequency step, f_s/N , i.e. (Heinzel et al., 2002):

$$PS(f_m) = \frac{k |y_m|^2}{N^2}; m = 0, 1, 2, \dots, \frac{N}{2} \quad (S3)$$

In the above Eq. S3, k takes values as for Eq. S2. The power spectrum has units of power, i.e. squared units of the input data \mathbf{x} . It represents the power in each frequency band, but not normalized to the frequency bin. Integrating P_{xx} over all frequencies or, equivalently, summing the elements of PS results in the variance of \mathbf{x} . This relationship is known as Parseval's theorem and was verified by testing that the variance computed in the frequency domain matches the time domain variance (normalized by N) to within machine precision. This also verifies that the scaling of the DFT was correct (Heinzel *et al.*, 2002). Previous use of Fourier analysis for phenology analysis is documented, for example, in Moody and Johnson (2001). De Beurs and Henebry (2010) provide an overview of various phenology analysis methodologies and their advantages and disadvantages.

The annual seasonal cycle was modeled using the fundamental frequency $f = 1 \text{ yr}^{-1}$ and all its available harmonics, i.e. all frequency bands that are its integer multiples, up to the Nyquist frequency of 6 yr^{-1} ($f = 1, 2, \dots, 6 \text{ yr}^{-1}$). That is, the signal was modeled as the summation of sinusoids whose amplitudes and phases are determined by the Fourier coefficients y_m at the respective harmonic frequencies (MathWorks, 2015):

$$\hat{\mathbf{x}} = a_0 + a_n \cos(2\pi ft) - b_n \sin(2\pi ft); f = [1; 6], f \in \mathbb{Z} \quad (\text{S4})$$

In the above, $\hat{\mathbf{x}}$ represents the modeled signal, t represents time in years, a_0 corresponds to the signal mean and is equal to y_0 (Eq. S1). The a 's are the real part, and the b 's – the imaginary part – of the Fourier transform y_m at the corresponding harmonic frequencies, divided by N . Because of conjugate symmetry, the a and b coefficients were doubled for all frequencies except $f = 0 \text{ yr}^{-1}$ and f_{Nyquist} . Equation S4 above was applied at a fine temporal sampling of $\Delta t = 1/10^{\text{th}}$ of a day approximately. However, the time of maximum is aggregated to a month because the underlining data sets have monthly resolution.

Peak analysis is then applied to $\hat{\mathbf{x}}$ as described in Sect. 2.3, in order to determine the relevant phenological parameters. The percent variance explained by the modeled signal was calculated as the summation of the power spectrum (PS , Eq. S3) terms corresponding to $f = 1, 2, 3, 4, 5$, and 6 yr^{-1} , divided the total variance of the input data vector \mathbf{x} (Eq. S1). An example of the DFT-based seasonal cycle modeling and peak analysis is illustrated in Fig. S1.

As a verification of the DFT and peak analysis methodology used to derive phenological parameters, we applied the methodology to monthly SeaWiFS photosynthetically available radiation (PAR, $\text{mol photons m}^{-2} \text{ day}^{-1}$). We used PAR data because it is expected to exhibit strong predictable seasonality in most of the world. For brevity, PAR analysis results are not plotted here, but results are summarized as follows. The derived seasonal amplitude was low near the Equator and higher poleward, as expected. The percent variance explained by the DFT-modeled signal indicated that in most of the world PAR seasonality is well captured by a sinusoidal model and explains most of the variability of the signal. The month of maximum for the primary peak in monthly PAR is December or January in most of the Southern Hemisphere and June, July or August in most of the North Hemisphere. Equatorial locations are expected to have two peaks each year near the equinoxes, but the peaks need not be the same amplitude, so the primary peak occurs in March in some places and August or September – in others. The PAR primary peak duration tends to be highest near to, but not quite at, the Equator (where there are double peaks) and gets progressively shorter at higher latitudes in both hemispheres.

Since missing data (especially at high latitudes) can affect phenological analysis (*Sapiano et al., 2012; Cole et al., 2012*), we investigated the effect of data gaps on the DFT analysis used here. We computed monthly averages of top of the atmosphere (TOA) total solar irradiance (insolation) using the model of *Kostadinov and Gilb (2014)*. Since monthly TOA insolation (and thus PAR) can be exactly 0 W m⁻² for some months near the solstices during the corresponding polar nights, the same phenological analysis (Sect. 2.3 and S1) applied to the PFT data sets was also applied to two variants of the monthly TOA data set – the original data set with exact zeros near the Poles, and with those zeros replaced by missing data and interpolated over (as done with gaps in the PFT data sets). When interpolation was used and the zeros were ignored, the signal mean was increased and the amplitude was decreased by small amounts for latitudes above 70°. These effects are small for TOA insolation but could be more significant and could occur at lower latitudes for PAR and especially for PFT and Chl data. Interpolation over missing data can also introduce artificial non-sinusoidal waveforms, which will cause artificial peaks in the DFT spectra and the modeled signal. Observed small decreases of PAR amplitude poleward of ~45° (not shown) could therefore be due to real geophysical phenomena (such as cloudiness patterns), but they may also be modeling artifacts (missing PAR data starts to occur for at least one month per year at these latitudes). Thus, results of analysis presented here at high latitudes or in other areas of frequent missing data must be interpreted with caution (see also Fig. S6).

Part 2. IPCC Models (CMIP5) Data and Phenological Parameters

In addition to analyzing phenology from various ocean color PFT algorithms (Table 1), we also investigated the same phenological parameters in a group of Earth System simulations from the recent Coupled Model Inter-comparison Project CMIP5 (*Taylor et al., 2012*). CMIP5 model output was downloaded from <http://pcmdi9.llnl.gov/esgf-web-fe/>. We derived phenological parameters from the same five years of “present” historical output (2003 to 2007) of the variable ‘phydiat’ (“mole concentration of diatoms expressed as carbon in seawater”). Diatom carbon concentration was chosen because it is most similar to the large phytoplankton variables of the satellite algorithms (Table 1); hence only those models that provide this variable are used. The “present” output for 2003-2005 is based on the historical scenario (years 1850 to 2005) forced by observed atmospheric changes (both anthropogenic and natural). The last two years (2006 and 2007) of the “present” output are based on the RCP8.5 scenario (*Riahi et al., 2011*). Table S1 provides details and references for the models. CMIP5 model details and global analysis of ecology for the present and for the 21st century for this same subset of models was presented in *Cabré et al. (2015)*. Molar concentration provided by the models (mol diatom C m⁻³) was converted to mass concentration (mg C m⁻³) using the atomic weight of carbon (12.011 g/mol, *Wieser et al., 2013*). All model output was resampled to a 1° grid before applying the DFT calculations (Sect. 2.3). Before computing phenological parameters, biomass values below 0 were set to missing data.

Table S1. Summary of the CMIP5 models that were used in the phenology inter-comparison study presented here. These models were selected because they explicitly model diatom biomass.

<i>Model</i>	<i>Nutrients</i>	<i>Ecology module</i>	<i>Phytoplankton variables</i>	<i>References</i>
CESM1-BGC	P, N, Fe, Si	MET	Diatoms, small phytoplankton, diazotrophs	Moore et al. (2004), Moore et al. (2006)
GFDL-ESM2G (M)	P, N, Fe, Si	TOPAZ2	Large phytoplankton (diatoms, greens, and other large eukaryotes), small phytoplankton (prokaryotic picoplankton and nanoplankton), and diazotrophs	Dunne et al. (2013)
HadGEM2-ES	N, Fe, Si	Diat-HadOCC (NPZD)	Diatoms, non-diatoms	Palmer and Totterdell (2001)
IPSL-CM5A-MR	P, N, Fe, Si	PISCES (from HAMOCC5)	Diatoms, nanophytoplankton (non-diatom).	Aumont and Bopp (2006), S��f��rian et al. (2013)
GISS-E2-H-CC (GISS-E2-R-CC)	N, Fe, Si	NOBM	Diatoms, chlorophytes, cyanobacteria, coccolitophores	Gregg (2008)

Part 3. Statistics of Circular Quantities

Quantities such as day or month of maximum are circular and can be transformed to an angular representation, i.e. angles close to 0 are also close to 2π , and December is temporally close to January. Thus, conventional descriptive statistics are not applicable to such quantities, e.g. for the calculation of the mean and variance of a set of months or angles (*Berens, 2009*). In particular, differences between two months cannot be simply calculated as it is done for linear differences. To resolve this issue and provide meaningful difference and mean and variance estimates of circular quantities used here, we employed circular statistics principles. The signed difference between two months was computed by first representing months in angular form (each month spans 30 degrees, mid-January is at 15° , mid-February at 45° , etc.). These angles were treated as vectors on the units circle and the absolute value of the angle between them was calculated using their dot product. This angle was converted to time units and represented the time difference between two months. The sign of the difference was determined by the sign of the sine of the angle resulting from subtracting the angular representations of the two months. This sign indicates which value leads and which lags in time. By convention, the difference is positive when the first element of the subtraction leads in time.

To calculate the mean month of maximum among the algorithms or models, the months were also represented in angular form and the *i* and *j* components of the corresponding vectors were averaged separately. The resultant vector's orientation, calculated with the arctangent function, determined the mean months of maximum. This method is consistent with *Berens (2009)*. Variance was computed as one minus the length of this resultant vector (*Allen and Johnson, 1991; Berens, 2009*). Unlike the variance of linear quantities, circular variance is bound between 0 and 1.

Part 4. Methodology for Regionally Binned Analyses (North Atlantic)

In addition to the per-pixel analysis of phenology, satellite algorithm and model data were spatially binned and regional analysis was performed for two example *Longhurst (1998)* marine biogeographical provinces (obtained from the Flanders Marine Institute (*Claus et al.*, 2013)), representing regimes of 1) North Atlantic spring bloom region (Westerlies - North Atlantic Drift – NADR, 3.512×10^6 km²), and 2) a subtropical gyre region (North Atlantic Subtropical Gyre West - NASW, 5.809×10^6 km²). Data for the spatial averages were extracted from the 9-km original imagery for the satellite data (PhytoDOAS data were first resampled to 9 km using nearest neighbor interpolation) and from the 1-degree data for the CMIP5 models. The scalar regional value was obtained by calculating the weighted pixel average, the weights being pixel area. Chl-based data (OC4v6 Chl and PhytoDOAS, Table 1) were averaged in log space. If zeros were present within the region, those pixels were excluded from the log-space averaging, but the final spatial average was weighted by the number of non-zero pixels divided by the number of valid pixels (*Habib, 2012*). Monthly climatological time series were also computed for the Longhurst provinces by averaging the data for all years of a given month (2003 to 2007) after the spatial binning. The same DFT-based phenological analysis was performed on the regionally binned complete time series (not the monthly climatology) as for the per pixel analysis. (Sect. 2.3, Supplement Sect. S1). The PHYSAT algorithm frequency of diatom detection variable (Table 1) is characterized by a large degree of sparsity and exact zeros that render the per-pixel DFT-based phenology analysis impossible. Regionally binned analysis addresses this sparsity issue. In order to link the NASW satellite data analysis to at-sea observations, Bermuda Atlantic Time Series (BATS) *in situ* primary production and sediment trap data were downloaded from <http://bats.bios.edu>. Details of BATS *in situ* methodology are given in the caption of Fig. S11.

Part 5. Details on Month of Maximum

Several algorithms that closely resemble Chl also exhibit minimal differences with the ensemble mean month of maximum (Fig. S2), especially BR10, OC-PFT, CB06, and to a lesser extent UITZ06 and MY10. This result is not surprising for the abundance-based algorithms (BR10, OC-PFT, UITZ06) (Table 1) as these are parameterized with Chl; thus their retrievals are a strong function of Chl. For spectral-based models (CB06, MY10), this suggests the temporal variations of derived spectral absorption properties of phytoplankton are generally consistent with those of Chl (i.e. the absorption spectrum becomes flatter, indicating, as generally expected, an increasing proportion of microplankton when Chl concentration increases). In contrast, the remaining algorithms exhibit larger excursions from the ensemble mean, with the biggest differences occurring in the Southern Ocean for KSM09 and ROY13 (both exhibiting a similar pattern), the subtropical gyres for FUJ11, and many widespread areas for ROY13 and PhytoDOAS. Differences are expected for FUJ11 in the gyres as this algorithm is developed with high-latitude regional data only. PhytoDOAS differences may be attributable to some extent to the use of data from a different sensor with a different spatial and temporal resolution, and retrieving diatom Chl and not microphytoplankton fraction. It is not clear how to explain the differences in ROY13. Overall, since secondary blooms are also detected (Sects. 2.3 and 3.5; this Supplement Part 7), it is possible that some secondary and primary blooms are hard to distinguish and can be confused due to data noise, if their amplitudes are similar. This is particularly true when dealing with fractional biomass, as opposed to absolute units (*Cabré et al.*, 2016). This similarity in amplitude is particularly evident in KSM09 data in the temperate

and subarctic North Atlantic (e.g. Fig. S1). Note that since KSM09 is based on backscattering, it may be detecting phases of the bloom differently, e.g. because peak carbon may not be the same as peak Chl.

Part 6. Details on Primary Bloom Duration

Maps of primary bloom duration for Chl and the individual PFT algorithms (Fig. S4) reveal significant differences among them, as well as significant variability of high spatial frequency (likely noise). The spatial patterns for Chl determined with the DFT analysis here closely agree with the observations of *Racault et al. (2012)* (see their Fig. 1d). The Chl spatial patterns of Fig. S4 (as well as the ensemble mean PFT-based patterns of Fig. 4A) also agree to first order with the results of *Sapiano et al. (2012)*. Namely, *Sapiano et al. (2012)* also observe zonal maxima at $\sim 30^\circ$ N/S latitude as the most prominent feature of the bloom duration global map. Durations there are quantified at ~ 170 -180 days (see their Fig. 9), roughly consistent with the long bloom durations (> 120 days, often up to 180 days) as quantified here by the DFT method for Chl (Fig. S4, top left panel). Other features are also in agreement, such as the long bloom durations equatorward of the subtropical gyres, in the entire South Atlantic gyre, as well as in parts of the subarctic Atlantic. As a quantification of the level of agreement in terms of bloom duration, the standard deviations of bloom duration for the satellite algorithms are depicted in Fig. S5A, and for the CMIP5 models – in Fig. S5B.

As with other phenology metrics, the abundance-based PFT algorithms are most similar to Chl (BR10 and OC-PFT in particular, and UITZ06 to a lesser extent). CB06 is also similar, indicating consistency between the spectral shape of absorption coefficients and Chl concentration. ROY13 and KSM09 exhibit long bloom durations in the Southern Ocean, unlike the rest of the algorithms. PhytoDOAS, MY10, and FUJ11 exhibit significantly shorter bloom durations overall globally, as compared to the other algorithms. Overall, there is considerable disagreement among the PFT algorithms with regards to the primary bloom duration metric. The standard deviation of bloom duration among the 10 PFT algorithms tends to be about 20 – 40 days over much of the ocean, but it can reach up to 70 days in certain areas (Fig. S5A). The 7 CMIP5 models also exhibit high standard deviation of primary bloom duration (Fig. S5B); however, most high values occur in the tropics and subtropics, whereas higher latitudes generally exhibit better model agreement with regards to bloom duration (but this observation could be influenced by missing data at high latitudes, Fig. S6).

Part 7. Details on Secondary Blooms

Note that the maps of Fig. 5A and Fig. 5B are not exactly complementary to each other (one is not equal to unity minus the other) as some areas exhibit a complex signal with more than two peaks which may or may not be ecologically significant (they are not analyzed here). The model of *Sapiano et al. (2012)* detects a double peak in SeaWiFS Chl data in the Pacific at 40° N, which is stranded by a single peak zone to the south and a flat model (no peak) to the north. This is roughly consistent with the PFT-based analysis presented here, except that the flat zone is identified as a single peak zone, and the PFT data has high data sparsity further north. The subtropical gyre and the Equatorial region in the Pacific exhibit many pixels of double peaks according to most PFT algorithms in the North Pacific, unlike *Sapiano et al. (2012)*, who identify a flat seasonal cycle there. Note that *Sapiano et al. (2012)* choose among 8 different

models (including sinusoidal and flat models, and including a secular trend), which is a different methodology from the analysis employed here.

Additional quantitative description of the secondary bloom can be provided by the phase difference between it and the primary bloom, as well as the relative strength of the secondary bloom compared to the primary one. The phase difference between the primary and secondary bloom, i.e. their relative timing, is given here as the difference in months between their peak timing (Fig. S7). The mean fractional prominence of the secondary blooms across the 10 PFT algorithms (Fig. S8A) indicates that in most of the areas where secondary blooms tend to occur (Fig. 5B) fractional prominence tends to be between 30 % and 60%. The equatorial secondary blooms tend to be of high fractional prominence, suggesting two annual blooms of roughly equal strength. The CMIP5 model's mean fractional prominence has a very different spatial pattern (Fig. S8B), again emphasizing a lack of the secondary bloom at temperate latitudes in the models. In comparison to the mean of the 10 PFT algorithms, Chl phenology exhibits fewer places with a secondary bloom (Fig. S8C). Importantly, the North temperate Atlantic area that does have Chl secondary blooms, exhibits smaller fractional prominence than the corresponding satellite algorithm microplankton ensemble mean (cf. Fig. S8A). This is most likely due to the fact that fractional microplankton for most PFT algorithms is compared to absolute Chl units. There is evidence that secondary blooms tend to be more dominated by larger phytoplankton than the corresponding primary bloom (*Sommer, 1996; Cabré et al., 2016*), so the secondary blooms appear *more pronounced* (i.e. of higher relative prominence) if they are expressed in terms of microplankton fraction. Thus, phenology analysis may be different if absolute biomass or Chl is used in the analysis instead (*Cabré et al., 2016*).

Part 8. Details on Regionally Binned and BATS Time Series Analysis.

A regime equivalent to the Atlantic transitional subpolar-subtropical regime discussed in Sect. 3.6 exists in the North Pacific as well as in the Southern Ocean at 35-50°S. Bimodal peaks result from light limitation in winter, growth in spring, then nutrient limitation in the summer and growth in the fall. The dynamics of the North Atlantic transitional and subpolar regimes is explained by *Evans and Parslow (1985)* and agrees with the Sverdrup critical depth theory (*Sverdrup, 1953*). Deep wintertime mixing ensures light limitation and little production and zooplankton population in winter, despite high nutrient supply. High nutrients from the winter and slow recovery of zooplankton in the spring and/or a large zooplankton class that does not respond fast enough to growing phytoplankton populations allow a spring phytoplankton bloom of large diatoms or *Phaeocystis*. This bloom is terminated by a drop in nutrients and zooplankton grazing in the summer; the fall bloom starts when mixing re-introduces nutrients to the upper layer. The seasonally varying *Northern subtropics* regime centered around 30°N, is characterized, just like in the Pacific, by a single annual peak in winter or early spring and high seasonal variability. The NASW province and the BATS station (Fig. S9) discussed below are part of this regime; the Pacific Ocean shows an equivalent regime. Previous work has determined that this peak is due to seasonal entrainment of nitrogen into the mixed layer.

Fig. S11 shows a complementary analysis of satellite and *in situ* time-series for the nutrient-limited, subtropical BATS station located inside the NASW region (Fig. S9). Most algorithms (and CMIP5 models) show a single biomass spring peak in Feb - March, as expected from *in situ* observations in this well-studied region (for a review see *Lomas et al. 2013*). The climatologies

of biomass and Chl are similar at BATS and NASW (compare Fig. 6B and Fig. S11A). As expected, the algorithms show slightly more consistency in the timing of the spring peak at BATS compared to the larger NASW region. Mixed layers at BATS vary from 10 m in the summer to 100 - 400 m in the winter, depending on the strength and phase of the North Atlantic Oscillation (Bates, 2012). The spring peak in biomass is driven by vertical mixing bringing macronutrients into the euphotic zone during winter. This nutrient flux supports a short spring-bloom period of higher primary production (Fig. S11C, bottom) and enhanced chlorophyll and carbon in most taxa present (e.g. Goericke, 1998). The ROY13, FUJI11 and PhytoDOAS are, as in the NASW region case, the most anomalous compared to the mid-February biomass peak date in the algorithm mean (Fig S11B).

Part 9. Details on Sources of Uncertainty

Some additional disadvantages of the DFT technique apart from the ones discussed in Sect. 3.7) include the difficulty in providing confidence intervals, and (if the Fast Fourier Transform (FFT) algorithm is used for DFT computation, as it is here), the necessity for data to be sampled at regular intervals (i.e. necessitating interpolation over missing data and padding missing data at the edges of time series with zeros). In the high latitudes, interpolating over many winter months with missing data may dampen the amplitude and create artificial waveforms that appear as spurious energy in various frequencies (Cole *et al.*, 2012; Sapiano *et al.* 2012). This is, however, a problem intrinsic to satellite observations, not the DFT technique. Long time series are required to achieve good frequency resolution (deBeurs and Henebry, 2010). The DFT technique can only distinguish frequencies that are half the sampling frequencies, i.e. up to the Nyquist frequency. Note that limitations such as aliasing (discussed in Sect. 3.7) and the Nyquist frequency are fundamental theoretical limitations of information and signal theory, rather than specific drawbacks of the DFT technique.

Just like in the oligotrophic tropics and subtropics, there is evidence that nanoplankton may dominate the annual bloom in the Southern Ocean as well (Uitz *et al.*, 2006; Sadeghi *et al.* 2012). In the Southern Ocean, CMIP5 models and satellite data exhibit diverging relationships of total biomass to fractional biomass in different size classes (Cabré *et al.*, 2016). Large differences of the various phenological parameters among the PFT algorithms in the Southern Ocean indicate that satellite data there need to be treated with caution and algorithm parameterizations need to pay special attention to this critical region. For example, it has been shown that for the OC-PFT algorithm other parameterizations are necessary when the algorithm is used in the Southern Ocean (Soppa *et al.* 2014). Parameterizations are generally challenging due to the dearth of *in situ* data from this remote region, and due to the tendency to build globally applicable algorithms. The Southern Ocean is known to be an atypical region in terms of bio-optics (Uitz *et al.*, 2006) where satellite-derived Chl is generally underestimated (Kahru and Mitchell, 2010). Phytoplankton community structure could be one of the factors explaining this bias in ocean color Chl estimates (Sathyendranath *et al.*, 2001; Jonhson *et al.* 2013; Ward *et al.*, 2015), which itself is used to estimate PFTs in some algorithms. In addition, considerable contributions to backscattering by coccoliths (e.g. Balch *et al.*, 2005; Balch *et al.*, 2011) and bubbles (Zhang *et al.*, 2002; Randolph *et al.*, 2014) are known to occur in the Southern Ocean. Furthermore, due to the low sun angles and the polar night, as well as to considerable cloudiness, satellite data in the Southern Ocean is sparse and many algorithms do not have valid retrievals (Fig. S6), biasing spatial and temporal means. Thus there are several reasons for being cautious

when using satellite OCR-derived products in the Southern Ocean. For more details on the specificities of OCR in high latitudes, see *IOCCG (2015)*. Note that coccolithophores are fairly global in distribution and anomalous calcite backscattering can affect some algorithm results elsewhere, because it can introduce errors in band-ratio-derived Chl (*Balch et al., 2005*), or cause violations in the assumptions of the KSM09-based PFT retrievals, for example. In terms of phenology (specifically month of maximum), their confounding effect is likely mitigated to some degree depending on the algorithm, because calcite concentrations tends to co-vary with Chl (*Hopkins et al., 2015*). Also note that some of the algorithms specifically retrieve coccolithophores as a PFT (PHYSAT, PhytoDOAS), and thus take their effects into account, and NASA standard processing implements a high coccolithophore concentration flag, thus masking out some of their confounding signal.

Algorithm users need to keep in mind that algorithms (especially those based on empirical relationships) can only be as good as the data sets used to develop them; *in situ* data set uncertainty translates to algorithm failure or uncertainty. Development data sets are not necessarily representative of the ecosystem states geographically or temporally even within the regions they span. Finally, ensemble means may be biased toward algorithms based solely on Chl as input (3 of the 10 PFT algorithms are Chl-based).

Oceanic ecosystems are expected to exhibit biennial variability (*Platt et al., 2009*) and more complex interannual variability due to climate oscillations such as ENSO and PDO, as well as secular trends due to climate change. Thus next steps in the PFT phenology analysis need to involve longer time series and allow for temporal evolution of phenology, e.g. via wavelet analysis (or studying the fractional frequencies of the DFT, or using a sliding DFT), and allow for a secular trends in the model (*Weatherhead et al., 1998; Sapiano et al., 2012*). However, a longer time series is recommended for that. Such a time series needs to be self-consistent, which requires seamless merging of several successive satellite mission. Apart from phenology, study of long term trends using seamlessly merged satellite data sets (*Maritorena et al., 2010*) is important, focusing on a power analysis of trend estimation (*Gerrodette, 1987*), i.e. predicting the length of record required to distinguish trends from noise in the data (e.g. *Beaulieu et al., 2013; Kostadinov and Lookingbill, 2015*).

Additional Supplement References*

(*references appearing in the main text and also cited in the Supplement are not repeated here)

Allen, F. H., & Johnson, O. W. E. N. (1991). Automated conformational analysis from crystallographic data. 4. Statistical descriptors for a distribution of torsion angles. *Acta Crystallographica Section B: Structural Science*, 47(1), 62-67.

Aumont O, Bopp L (2006), Globalizing results from ocean in situ iron fertilization studies, *Global Biogeochemical Cycles* 20:15. doi:10.1029/2005gb002591.

Balch, W. M., Howard Gordon, B. C. Bowler, D. T. Drapeau, E. S. Booth. 2005. Calcium carbonate measurements in the surface global ocean based on MODIS Data. *Journal of Geophysical Research-Oceans*. 110, C07001 doi:10.1029/2004JC002560.

Balch WM, Drapeau DT, Bowler BC, Booth ES, Lyczkowski E, Alley D (2011) The contribution of coccolithophores to the optical and inorganic carbon budgets during the Southern Ocean Gas Experiment: New

evidence in support of the "Great Calcite Belt" hypothesis. *Journal of Geophysical Research-Oceans Special Issue*. VOL. 116, C00F06, doi:10.1029/2011JC006941.

Bates, N. R.: Multi-decadal uptake of carbon dioxide into subtropical mode water of the North Atlantic Ocean, *Biogeosciences*, 9, 2649-2659, doi:10.5194/bg-9-2649-2012, 2012.

Beaulieu, C., Henson, S. A., Sarmiento, Jorge L., Dunne, J. P., Doney, S. C., Rykaczewski, R. R., and Bopp, L. (2013), Factors challenging our ability to detect long-term trends in ocean chlorophyll, *Biogeosciences*, 10, 2711-2724, doi:10.5194/bg-10-2711-2013, 2013.

Berens, P. (2009), CircStat: a MATLAB toolbox for circular statistics, *J. Statistical Software*, 31(10), 1-21.
Cole, H., S. Henson, A. Martin, and A. Yool (2012), Mind the gap: The impact of missing data on the calculation of phytoplankton phenology metrics, *J. Geophys. Res.*, 117, C08030, doi:10.1029/2012JC008249.

Claus S., De Hauwere N., Vanhoorne B., Hernandez F., Mees J. (Flanders Marine Institute) (2013), Marineregions.org, Accessed at <http://www.marineregions.org> on 2013-05-01.

Cole, H., S. Henson, A. Martin, and A. Yool (2012), Mind the gap: The impact of missing data on the calculation of phytoplankton phenology metrics, *J. Geophys. Res.*, 117, C08030, doi:10.1029/2012JC008249.

de Beurs, K. M., & Henebry, G. M. (2010). Spatio-temporal statistical methods for modelling land surface phenology. In *Phenological research* (pp. 177-208). Springer Netherlands.
Gerrodette, T. (1987), A power analysis for detecting trends, *Ecology*, 1364-1372.

Goericke, R., (1998), Response of phytoplankton community structure and taxon-specific growth rates to seasonally varying physical forcing in the Sargasso Sea off Bermuda, *Limnology and Oceanography*, 43, 921-935.

Gregg WW (2008), Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model, *Journal of Marine Systems* 69:205-225. doi:10.1016/j.jmarsys.2006.02.015

Heinzel, G., A. Rudiger and R. Schilling (2002), Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new flat-top windows, Max-Planck-Institut fur Gravitationsphysik, (Albert-Einstein-Institut), Teilinstitut Hannover.

Hopkins, J., S. A. Henson, S. C. Painter, T. Tyrrell, and A. J. Poulton (2015), Phenological characteristics of global coccolithophore blooms, *Global Biogeochem. Cycles*, 29, 239–253, doi:10.1002/2014GB004919.

IOCCG (2015). Ocean Colour Remote Sensing in Polar Seas. Babin, M., Arrigo, K., Bélanger, S. and Forget, M-H. (eds.), IOCCG Report Series, No. 16, International Ocean Colour Coordinating Group, Dartmouth, Canada.

Kahru, M., and B. G. Mitchell (2010), Blending of ocean colour algorithms applied to the Southern Ocean, *Remote Sensing Letters*, 1(2), 119-124

Kostadinov, T. S. and Gilb, R. (2014), Earth Orbit v2.1: a 3-D visualization and analysis model of Earth's orbit, Milankovitch cycles and insolation, *Geosci. Model Dev.*, 7, 1051-1068, doi:10.5194/gmd-7-1051-2014.

Kostadinov, T. S. and T. R. Lookingbill (2015), Snow cover variability in a forest ecotone of the Oregon Cascades via MODIS Terra products, *Remote Sensing of Environment*, 164C, pp. 155-169, doi:10.1016/j.rse.2015.04.002.

Lomas, M. W., Bates, N. R., Johnson, R. J., Knap, A. H., Steinberg, D. K., & Carlson, C. A. (2013). Two decades and counting: 24-years of sustained open ocean biogeochemical measurements in the Sargasso Sea, *Deep Sea Research Part II: Topical Studies in Oceanography*, 93, 16-32.

Maritorena, S., d'Andon, O. H. F., Mangin, A., & Siegel, D. A. (2010), Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. *Remote Sensing of Environment*, 114(8), 1791-1804.

MathWorks(2015), Fast Fourier Transform (FFT), URL: <http://www.mathworks.com/help/matlab/math/fast-fourier-transform-fft.html>; Last access: Dec. 14, 2015.

Moody, A., & Johnson, D. M. (2001). Land-surface phenologies from AVHRR using the discrete Fourier transform. *Remote Sensing of Environment*, 75(3), 305-323.

Moore JK, Doney SC, Lindsay K (2004), Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model, *Global Biogeochemical Cycles*, 18. doi:10.1029/2004gb002220.

Moore JK, Doney SC, Lindsay K, Mahowald N, Michaels AF (2006) ,Nitrogen fixation amplifies the ocean biogeochemical response to decadal timescale variations in mineral dust deposition, *Tellus Series B-Chemical and Physical Meteorology*, 58:560-572. doi:10.1111/j.1600-0889.2006.00209.x.

Palmer JR, Totterdell IJ (2001) Production and export in a global ocean ecosystem model, *Deep-Sea Res Part I-Oceanogr Res Pap*, 48:1169-1198. doi:10.1016/s0967-0637(00)00080-7.

Randolph, K., H.M. Dierssen, M. Twardowski, A. Cifuentes Lorenzen, C.J. Zappa (2014), Optical measurements of small deeply-penetrating bubble populations generated by breaking waves in the Southern Ocean. *J. Geophys. Res. Oceans*, 119, doi:10.1002/2013JC009227.

Riahi K et al (2011) RCP 8.5—a scenario of comparatively high greenhouse gas emissions. *Clim Change*, 109:33–57. doi:10.1007/s10584-011-0149-y

Sathyendranath, S., Stuart, V., Cota, G., Maas, H., Platt, T. (2001), Remote sensing of phytoplankton pigments: a comparison of empirical and theoretical approaches, *International Journal of Remote Sensing*, 22, 249–273.
Sommer, U. (1996). Plankton ecology: The past two decades of progress, *Naturwissenschaften*, 1270 83, 293–301. doi:10.1007/s001140050291.

Soppa M. A., Hirata T., Silva B., Dinter T., Peeken I., Wiegmann S., Bracher A. (2014) Global retrieval of diatoms abundance based on phytoplankton pigments and satellite. *Remote Sensing* 6: 10089-10106.

Sverdrup, H. U. (1953). On Conditions for the Vernal Blooming of Phytoplankton, *J. Cons. int. Explor. Mer*, 18 (3): 287-295. doi:10.1093/icesjms/18.3.287.

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design, *Bulletin of the American Meteorological Society*, 93:485-498. doi:10.1175/bams-d-11-00094.1.

Ward B.A. (2015), Temperature-Related Changes in Phytoplankton Community Structure Are Restricted to Polar Waters. *PLoS ONE* 10(8): e0135581. doi:10.1371/journal.pone.0135581

Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Meng, X. L., Choi, D. S., Cheang, W. K., Keller, T., DeLuisi, J., Wuebbles, D. J., Kerr, J. B., Miller, A. J., Oltmans, S. J., and Frederick, J. E. (1998), Factors affecting the detection of trends: Statistical considerations and applications to environmental data, *J. Geophys. Res.-Atmos.*, 103, 17149–17161, doi:10.1029/98jd00995.

Wieser, M. E., Holden, N., Coplen, T. B., Böhlke, J. K., Berglund, M., Brand, W. A., De Bièvre, P., Gröning, M., Loss, R.D., Meija, J., Hirata, T., Prohaska, T., Schoenberg R., O'Connor, G., Walczyk, T., Yoneda, S. and Zhu, X. K. (2013), Atomic weights of the elements 2011 (IUPAC Technical Report). *Pure and Applied Chemistry*, 85(5), 1047-1078.

Zhang, X., Lewis, M., Lee, M., Johnson, B., & Korotaev, G. (2002). The volume scattering function of natural bubble populations. *Limnology and Oceanography*, 47(5), 1273-1282.

Supplement Figures

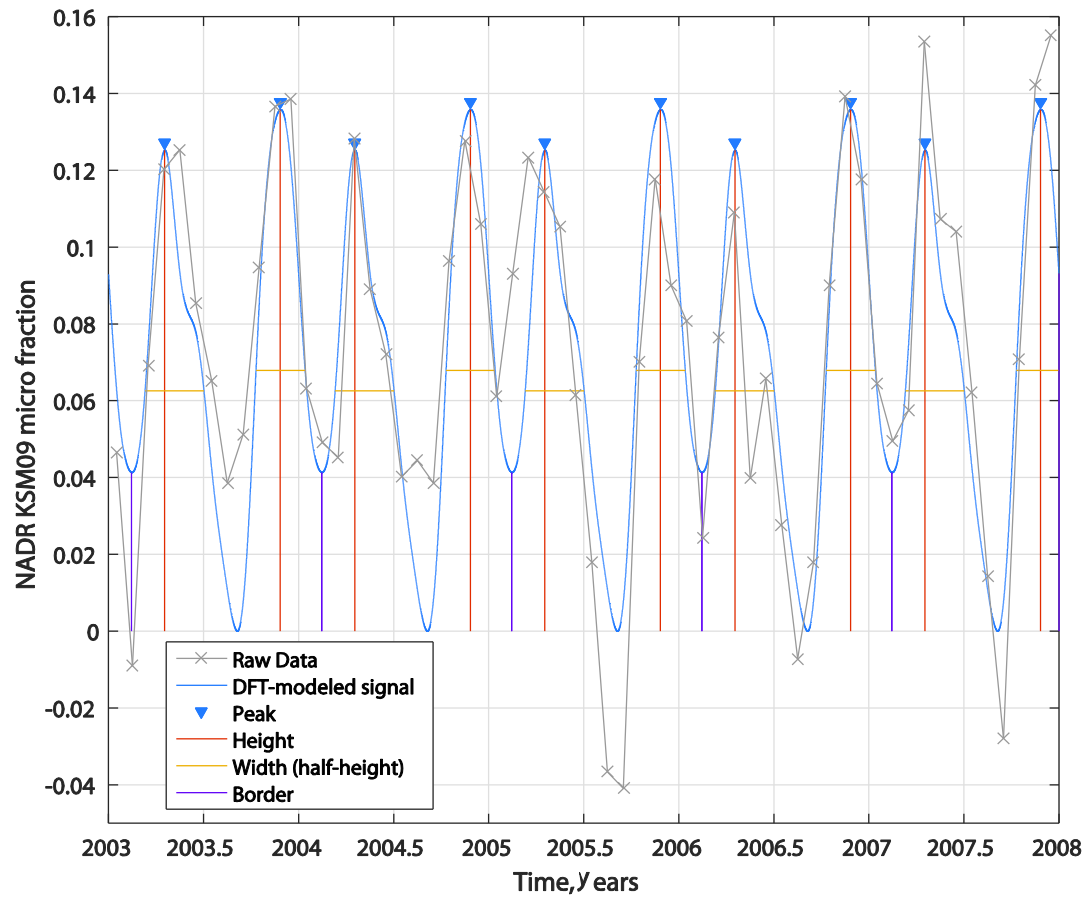


Figure S1. An example of DFT-based modeling of the annual seasonal cycle of PFT data, with subsequent peak analysis. The regionally binned KSM09 percent microplankton data for the *Longhurst (1998)* North Atlantic Drift province (NADR) is shown. The mean was subtracted from the data before modeling, and the minimum of the resulting modeled signal was subtracted before peak finding in order to identify height correctly. As a result, the modeled signal's minimum is zero, and some values of the raw data can be negative. Note the presence of two annual peaks of variable relative height in the data.

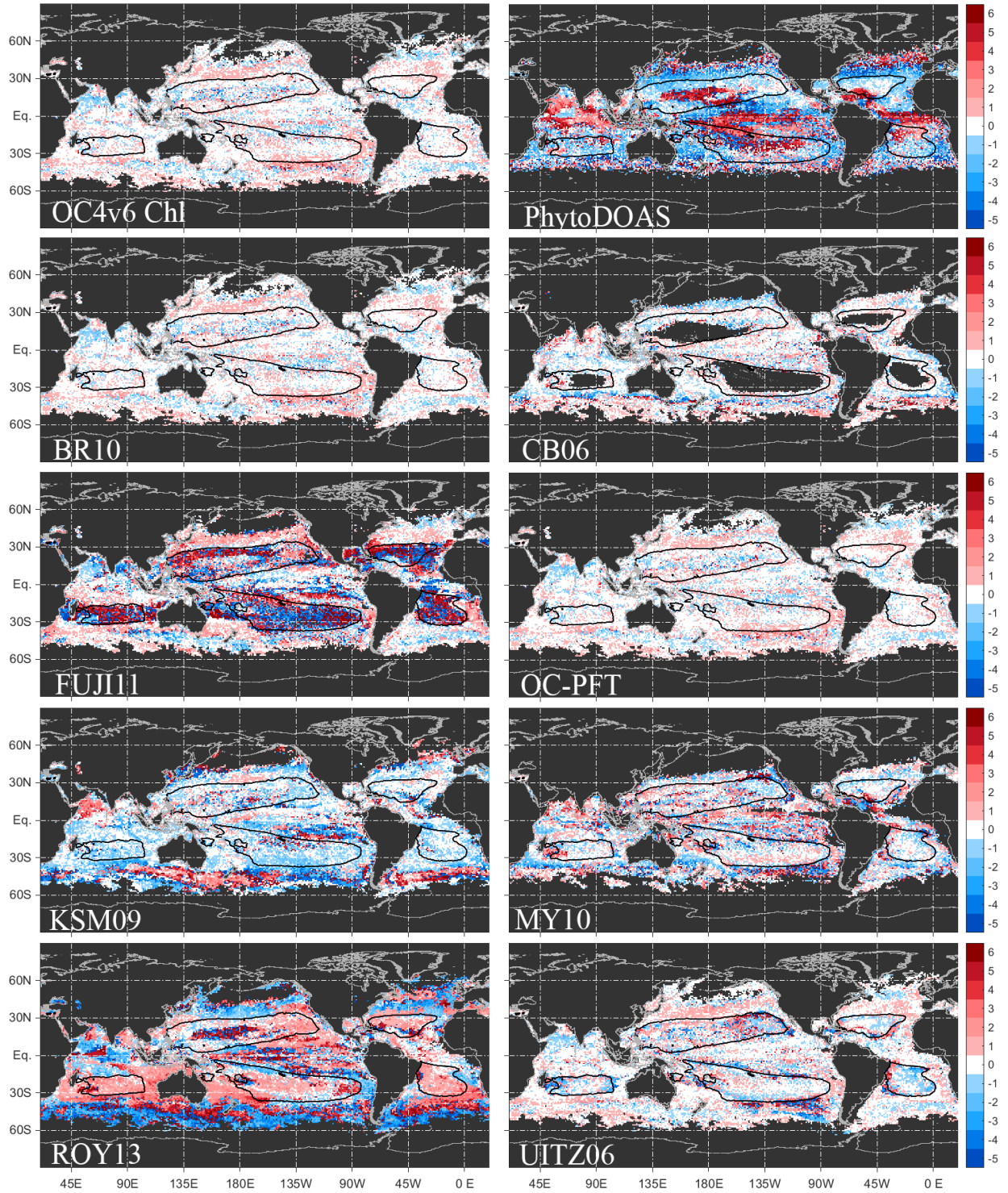


Figure S2. Maps of differences between the month of maximum bloom of OC4v6 Chl and the PFT algorithms (except PHYSAT, which exhibits very sparse data almost everywhere) and the ensemble mean of Fig 3A. Table 1 lists the algorithms and variables used. Positive differences indicate that the ensemble mean leads (i.e. occurs earlier than) the specific algorithm's month of maximum. Note that the determination of

month of maximum becomes unreliable when only a small percentage of the signal variance is explained by the seasonal cycle, thus these areas should be treated with caution (Figs. 2A, S13A). The isoline of climatological Chl = 0.08 mg m^{-3} is shown (black solid contour). Note that the CB06 algorithm retrieves almost no data in the subtropical gyres.

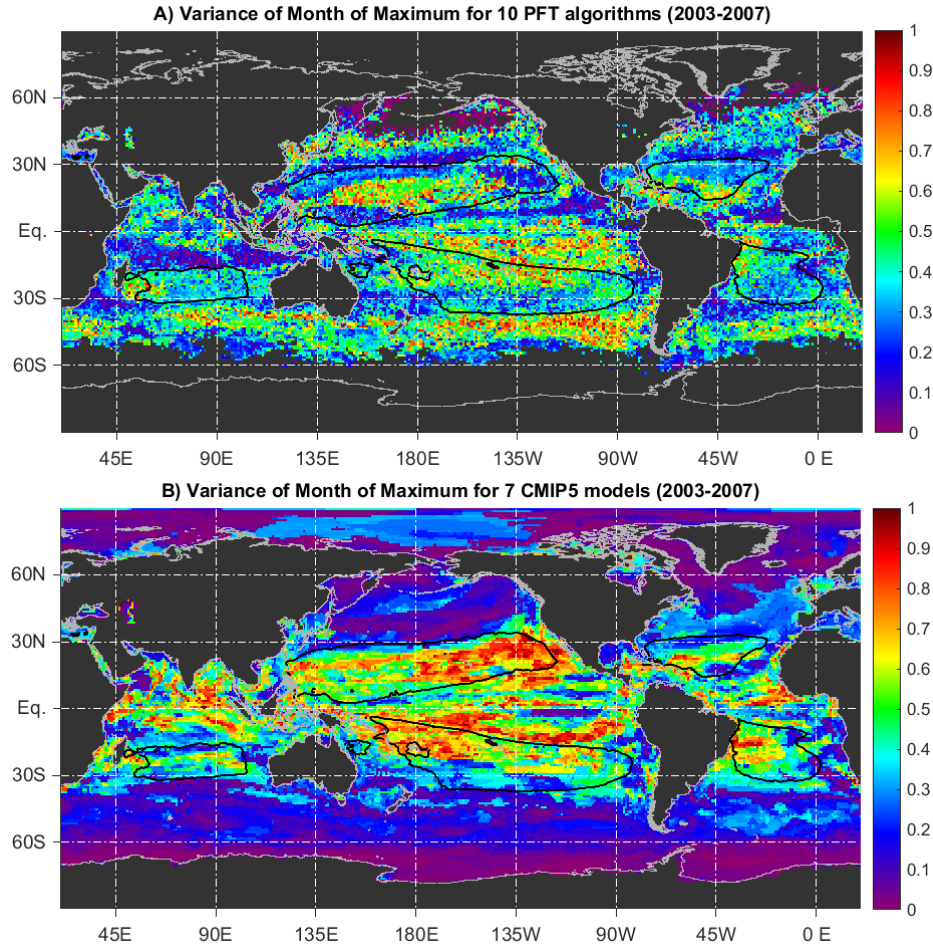


Figure S3. Circular variance of month of maximum of the primary bloom for (A) the 10 PFT algorithms and (B) the 7 CMIP5 models. The isoline of climatological Chl = 0.08 mg m⁻³ (black solid contour) is shown on both panels.

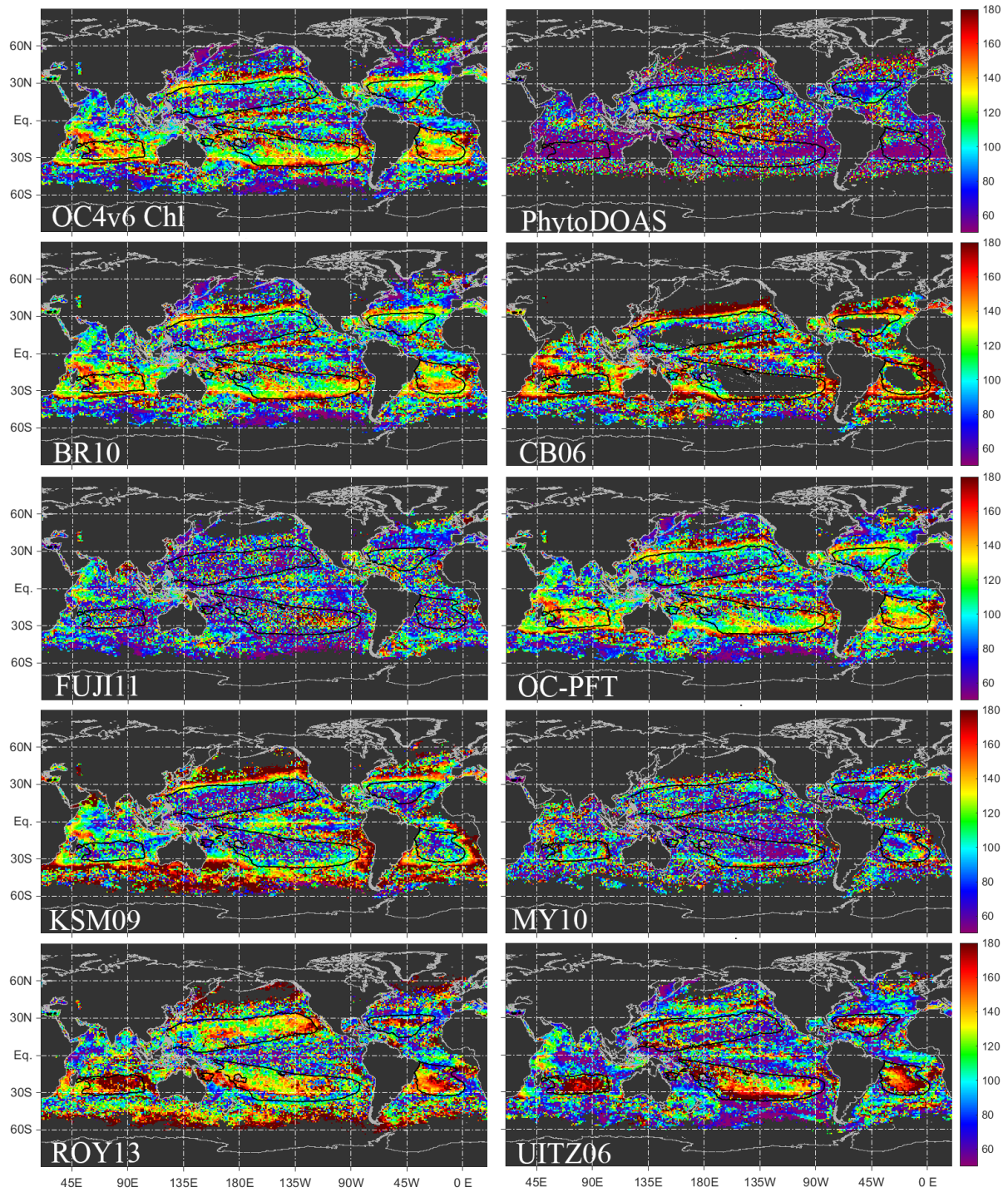


Figure S4. Maps of bloom duration (in days) for the primary bloom for OC4v6 Chl and the PFT algorithms (except PHYSAT). Duration of a bloom is defined as the width of the modeled seasonal signal at half the bloom peak height. The isoline of climatological Chl = 0.08 mg m⁻³ (black solid contour) is shown on all panels.

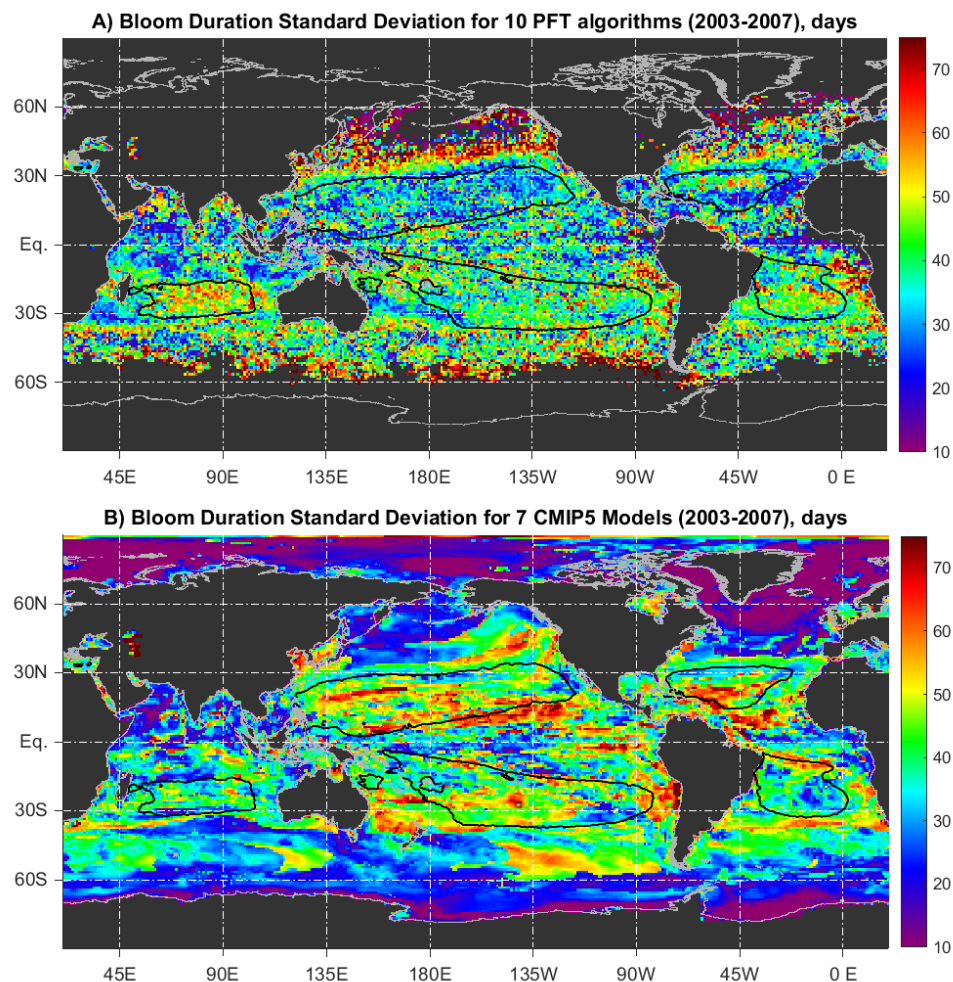


Figure S5. Standard deviation (in days) of the primary bloom duration for (A) the 10 PFT algorithms and (B) the 7 CMIP5 models. The isoline of climatological Chl = 0.08 mg m⁻³ (black solid contour) is shown on both panels.

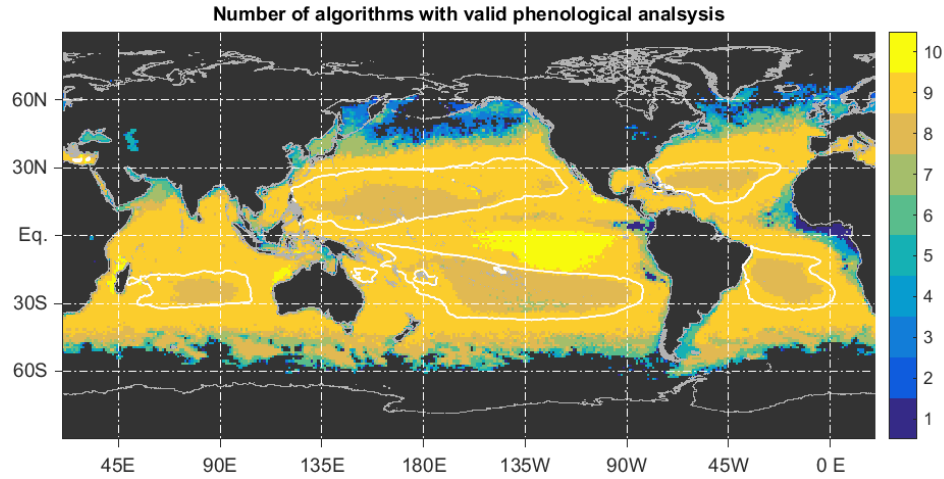


Figure S6. The number of algorithms for which valid phenological analysis is available. Compare with panels of Fig. 1, for example. The isoline of climatological Chl = 0.08 mg m⁻³ is shown as a white contour.

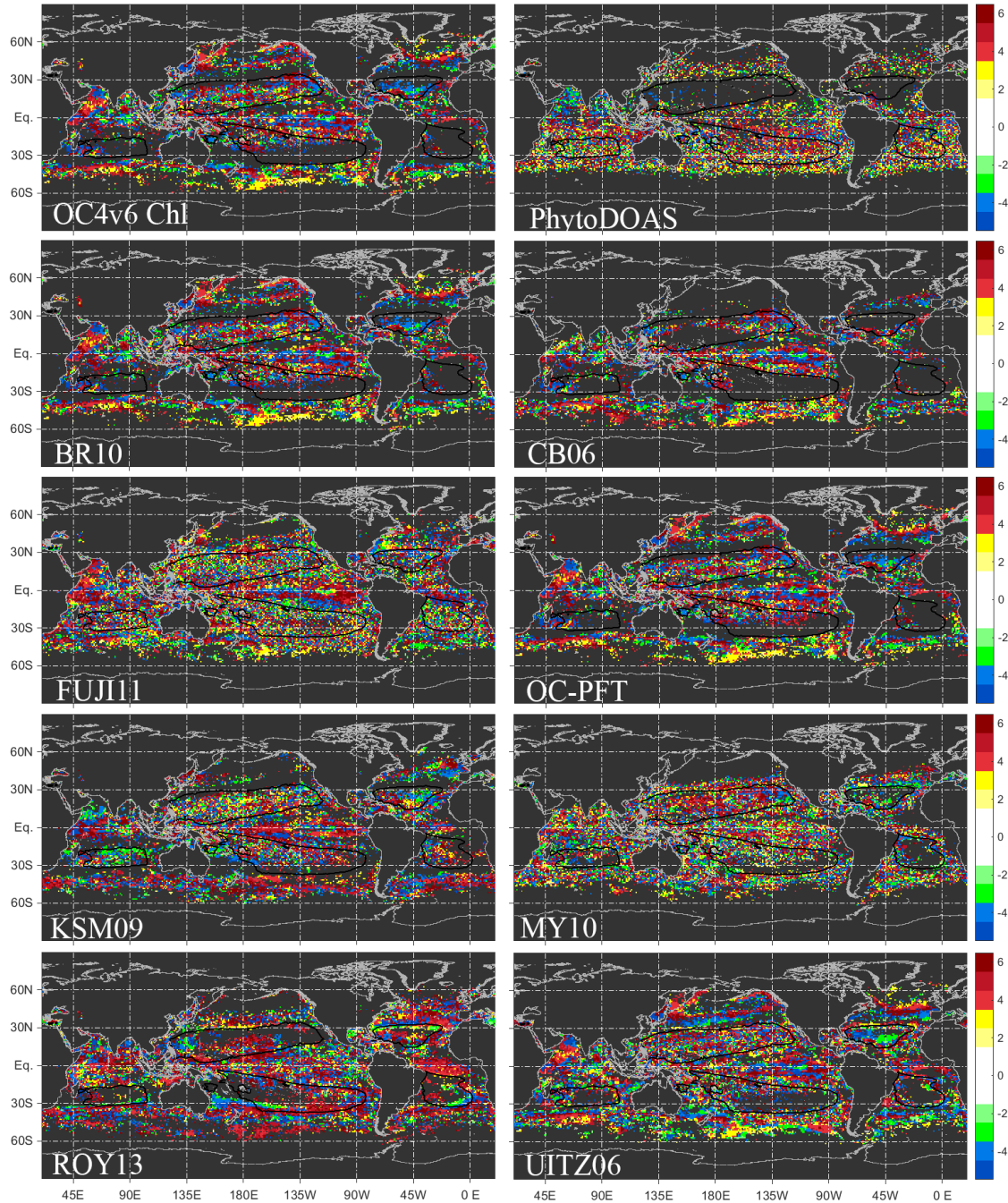


Figure S7. Difference in months between the maxima of the primary and the secondary blooms for OC4v6 Chl and the PFT algorithms (except PHYSAT). Differences greater than three months in absolute value are shown in either red colors for positive differences (the primary bloom leads, i.e. occurs earlier than the secondary bloom) or blue colors for negative differences (the secondary bloom occurs earlier than the primary bloom). A difference of six months is shown as positive by convention. The isoline of climatological Chl = 0.08 mg m^{-3} (black solid contour) is shown on all panels.

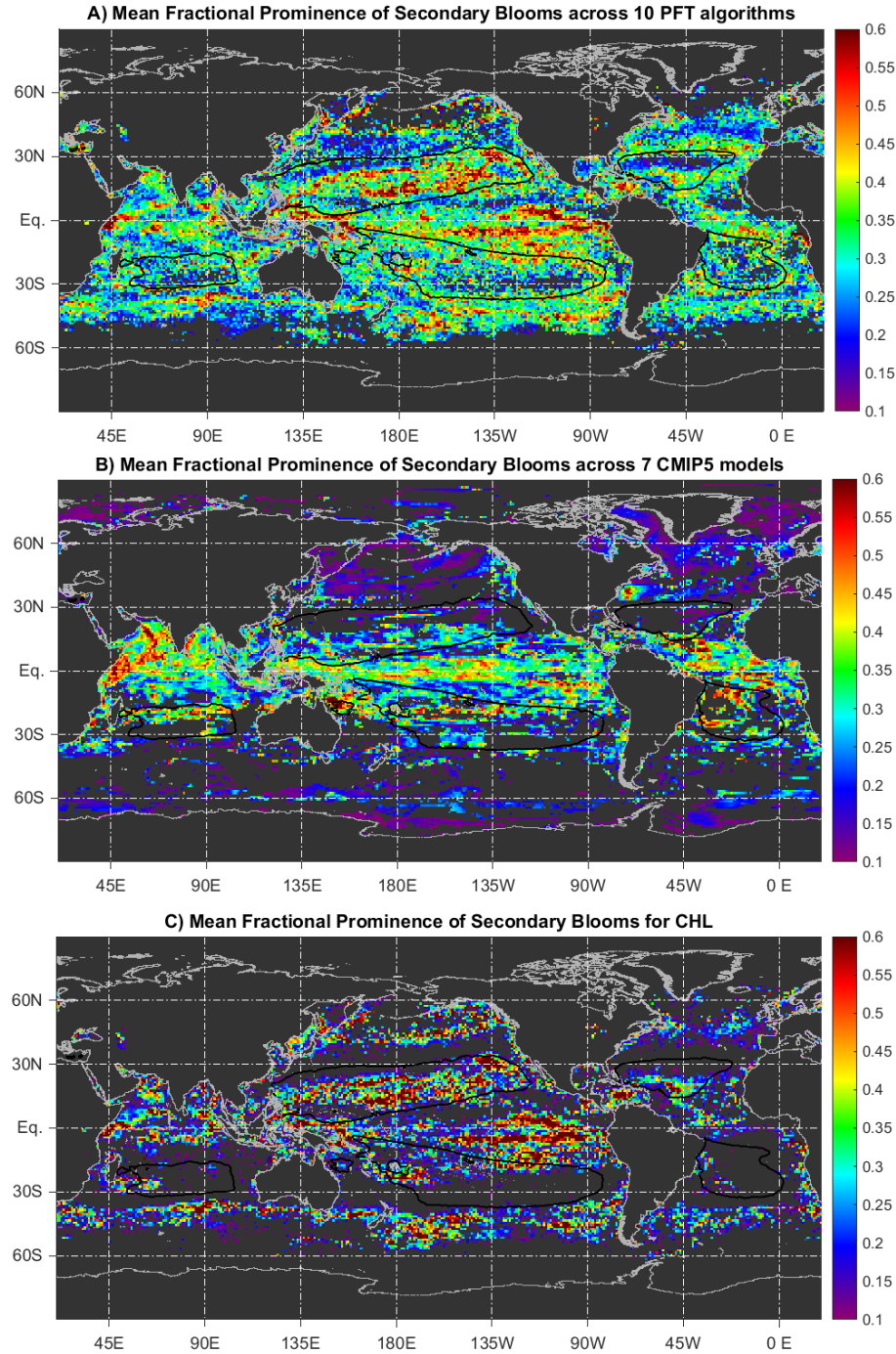


Figure S8. Fractional prominence of the secondary bloom in relation to the primary bloom, for, as follows: A) ensemble mean of percent microplankton for the 10 PFT algorithms, B) ensemble mean for diatom C for the 7 CMIP5 models, C) for OC4v6 Chl. Fractional prominence refers to the ratio of the prominence of the secondary bloom to the prominence of the primary bloom. The isoline of climatological Chl = 0.08 mg m^{-3} (black solid contour) is shown on all panels.

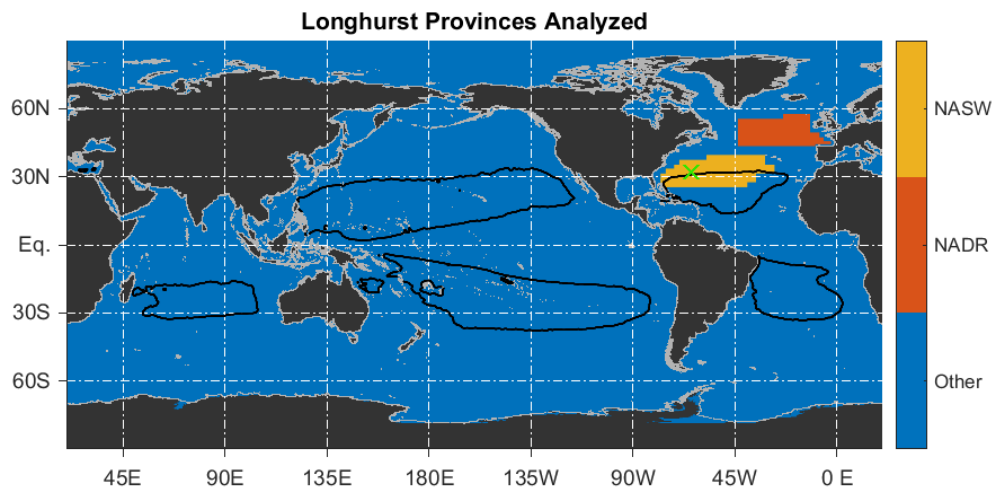


Figure S9. Map of the two *Longhurst (1998)* marine biogeographical provinces that were used for the regionally binned analysis – NADR (red) and NASW (orange). The location of the BATS station is indicated with a green cross (inside NASW). For details and province codes explanation, see this Supplement Part 4.

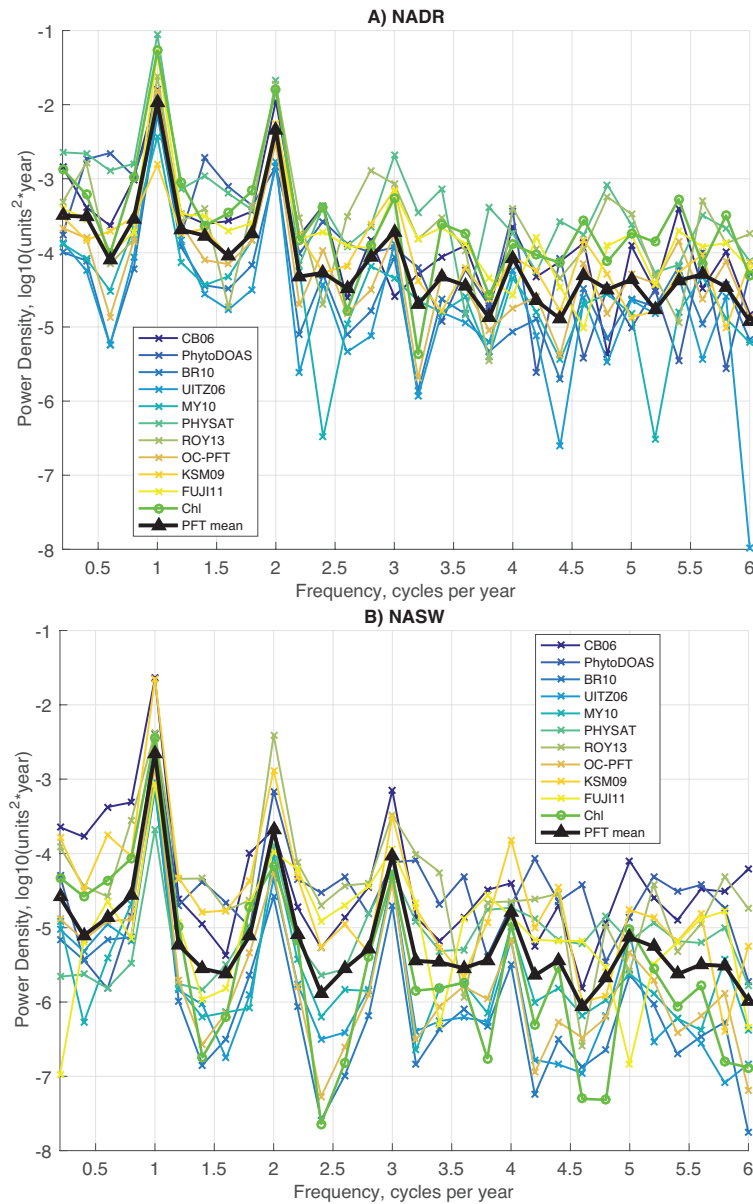


Figure S10. Power spectral density plots of the PFT algorithms (Table 1) and OC4v6 Chl for two example *Longhurst (1998)* biogeographic provinces as follows: A) North Atlantic Drift Region (NADR); B) The Western North Atlantic Subtropical Gyral Province (NASW). See Fig. S9 for a map of the provinces. See Sect. 2.3 and Supplement Sect. S1 for methodology details.

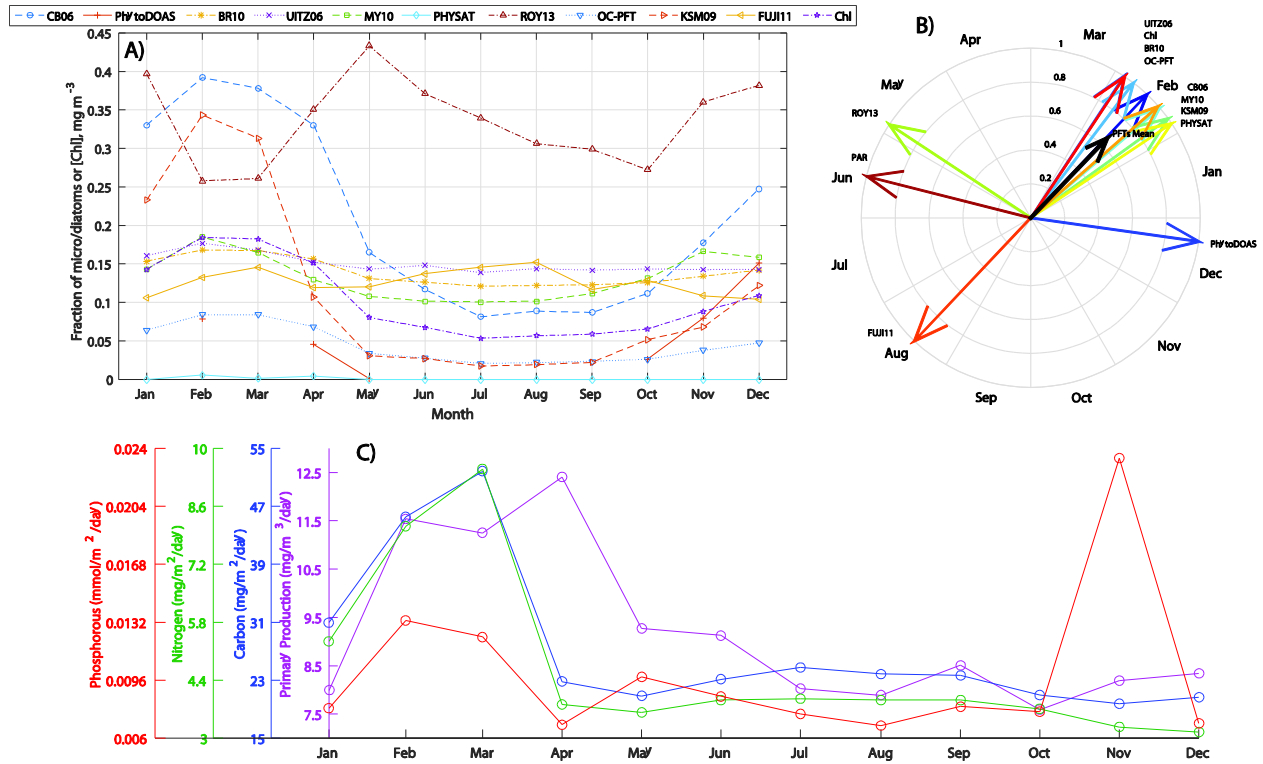


Figure S11. A) Time series as in Fig. 6, but for a 1x1° box centered on the BATS station only. B) As in Fig. 7, but for a 1x1° box centered on the BATS station only. C) Seasonal cycle of *in situ* biogeochemical data at the Bermuda Atlantic Time Series station at 31°50'N, 64°10'W. Purple: Primary Production, in mg m⁻³ day⁻¹. Go-Flo bottles were used to measure primary production at different depths. The shallowest depth for which there was consistent data, 5m, was used. Blue: Organic Carbon flux collected in sediment traps at the BATS station, in mg m⁻² day⁻¹. Sediment traps were left underwater at specific depths for 3 days, usually around once per month. 3 traps were used, and the average was taken. Again, the shallowest depth for which there was consistent data, 150m, was used. Drop sites for a given day varied by as much as half a degree. The traps would also float up to a degree during their 3-day deployment. Green: Nitrogen flux, in mg m⁻² day⁻¹, collected in the same traps as Carbon. Red: Phosphorous flux, in mmol m⁻² day⁻¹, collected in the same traps as Carbon. Climatologies for each variable were computed over the complete SeaWiFS period (September 1997 to December 2010).

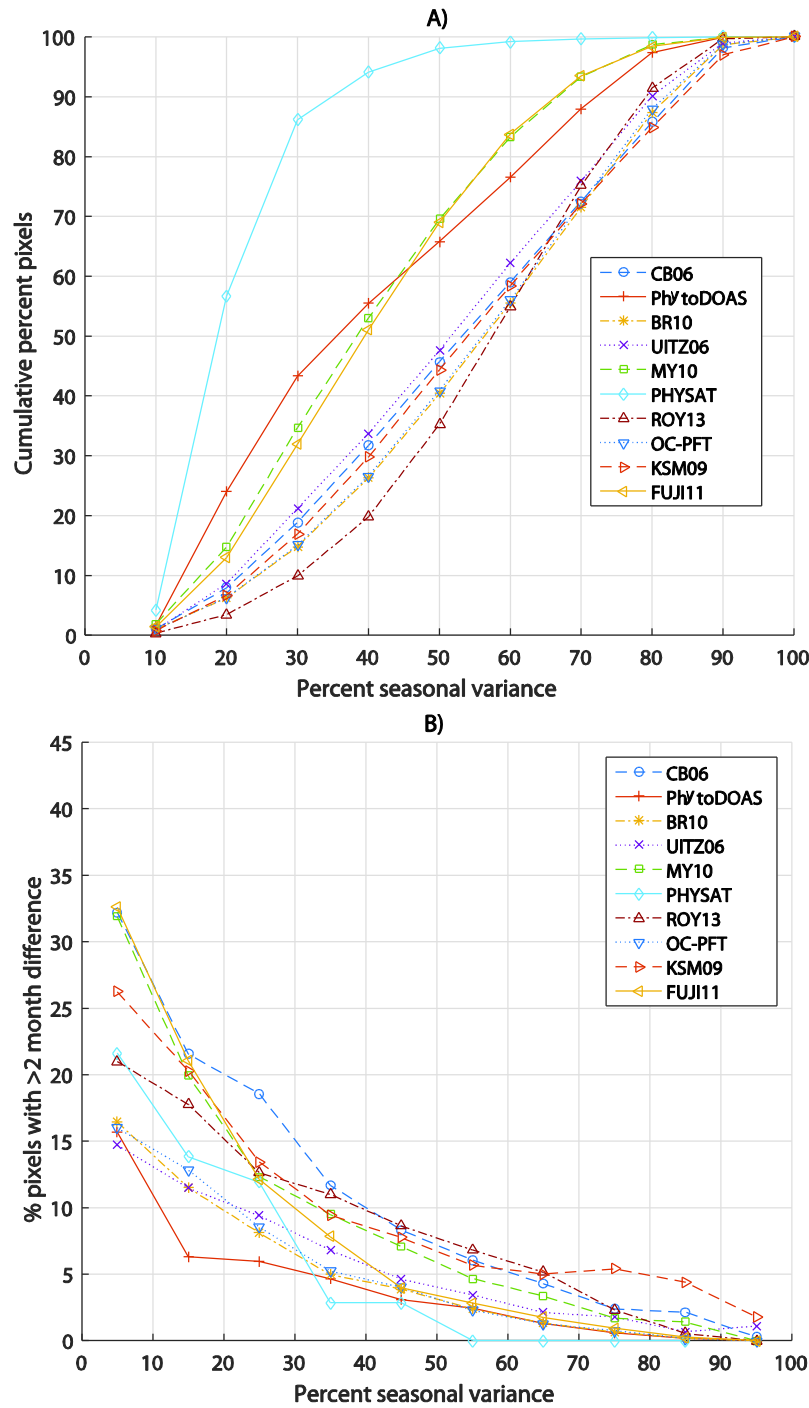


Figure S12. A) Cumulative percent of all pixels with valid phenological computations as a function of variance explained by the modeled seasonal cycle. For example, the graph indicates that for the MY10 model, ~70% of all analyzed pixels exhibit at most 50% variance explained by the modeled seasonal cycle. Note that differences among the algorithms can be intrinsic or can be due to differences in spatial coverage among the models (Fig. 1, Fig. S6), as illustrated prominently by the PHYSAT curve. B) Percentage of pixels exhibiting greater than 2 months difference in month of annual

maximum as determined from the DFT-based modeling used here, and as determined from direct peak analysis of the times series of the monthly climatologies of PFT algorithm satellite data (12 data points in each analyzed time series). The percentage is given as a function of percent variance explained by the DFT-modeled seasonal cycle, binned into 10%-bins. The percentage is calculated out of all pixels with valid phenology in each variance bin.

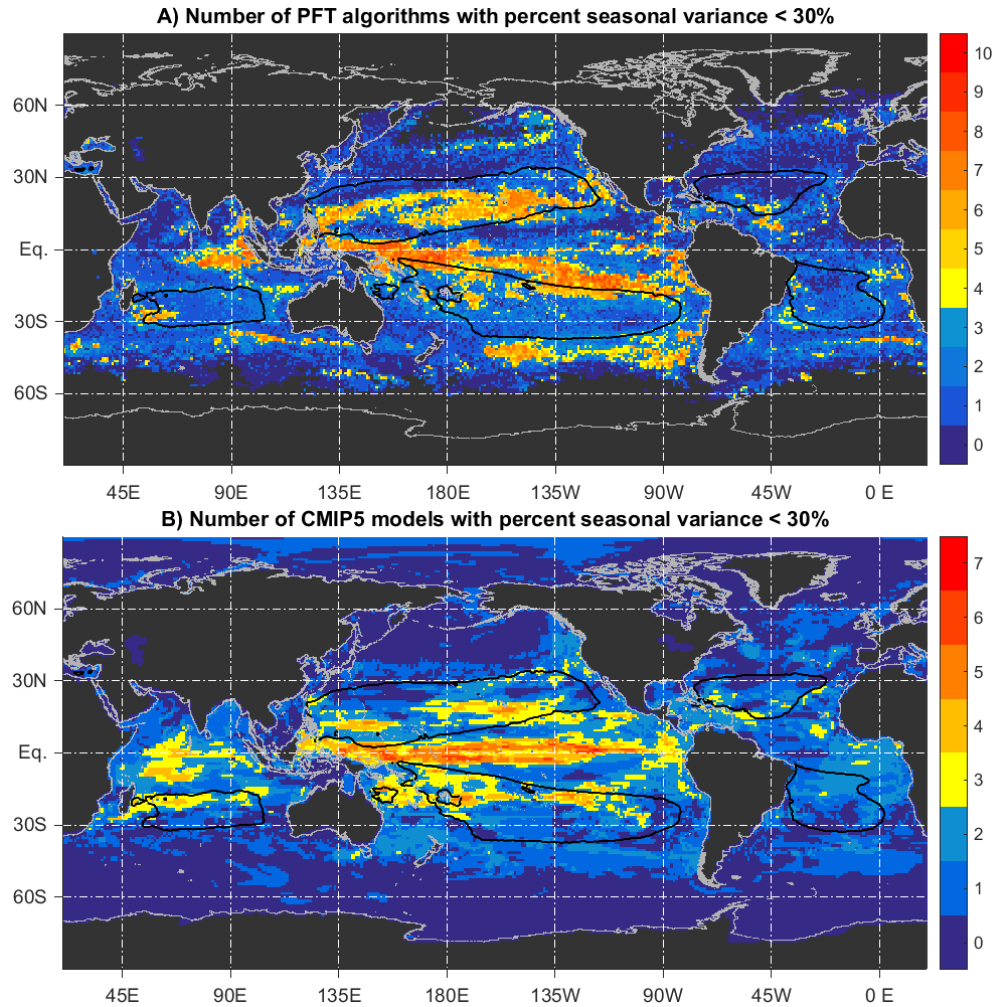


Figure S13. A) Number of PFT algorithms for which percent seasonal variance is less than 30%. Yellow and red colors indicate areas for which more than 3 PFT algorithms exhibit this, indicating that month of maxima (and other phenology metrics) should be interpreted with caution there. B) Same as A), but for the 7 CMIP5 models. Yellow and red colors indicate areas where more than two models exhibit less than 30% seasonal variance. The isoline of climatological Chl = 0.08 mg m^{-3} (black solid contour) is shown on both panels.